DESK

Volume 40 February 1954 Number 2

UNIVERSITY
OF MICHIGAN
APR 1 1954
PERIODICAL
READING ROOM

Proceedings of the

National Academy of Sciences

of the United States of America

The Proceedings of the National Academy of Sciences

OFFICERS OF THE ACADEMY

DETLEV W. BRONK
President

GEORGE W. CORNER
Vice President

ROGER ADAMS
Foreign Secretary

ALEXANDER WETMORE
Home Secretary

WILLIAM J. ROBBINS
Treasurer

EDITORIAL BOARD OF THE PROCEEDINGS

LINUS PAULING
Chairman

ALEXANDER WETMORE, *Home Secretary*
EDWIN B. WILSON, *Managing Editor*

ROGER ADAMS, *Foreign Secretary*
WILLIAM W. RUBEY, *Chairman of the National Research Council*

J. W. BEAMS
R. E. CLELAND
W. W. RUBEY
HARLOW SHAPLEY
F. E. TERMAN

G. W. BEADLE
E. A. DOISY
G. C. EVANS
C. H. GRAHAM
J. T. PATTERSON

GREGORY BREIT
L. C. DUNN
W. F. LIBBY
J. VON NEUMANN
PAUL WEISS

Subscription price is $7.50 per volume (12 issues); the price of a single issue is $.75. Make all remittances payable to The University of Chicago Press in U.S. currency or its equivalent. Authorized agent for the British Commonwealth, except North America and Australasia: Cambridge University Press, Bentley House, 200 Euston Road, London, N. W. 1, England.

Editorial correspondence should be addressed to the PROCEEDINGS of the NATIONAL ACADEMY of SCIENCES, Attention Miss Mary D. Alexander, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois.

Business correspondence should be addressed to the Journals Division, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois.

Subscribers are requested to notify The University of Chicago Press and their local postmaster immediately of change of address.

Microfilms of complete volumes of this journal are available to regular subscribers only and may be obtained at the end of the volume year from University Microfilms, 313 N. First Street, Ann Arbor, Michigan.

Entered as second-class matter June 21, 1920 at the Post-office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in Section 1103, Act of October 3, 1917. Authorized June 18, 1920.

PRINTED IN THE U. S. A.

THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES
is published for
THE NATIONAL ACADEMY OF SCIENCES
by
THE UNIVERSITY OF CHICAGO PRESS

N

T

L

Proceedings of the
NATIONAL ACADEMY OF SCIENCES

Volume 40 • Number 2 • February 15, 1954

CYCLICAL BEHAVIOR IN PNEUMOCOCCAL GROWTH AND TRANSFORMABILITY OCCASIONED BY ENVIRONMENTAL CHANGES

By Rollin D. Hotchkiss

LABORATORY OF PATHOLOGY AND MICROBIOLOGY, THE ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH, NEW YORK

Communicated by R. J. Dubos, December 23, 1953

The availability of a system in which drug sensitive bacteria are transformed to drug resistance through the action of desoxyribonucleate (DNA) extracted from resistant organisms[1] makes possible a quantitative determination of the number of bacteria transformed under given experimental conditions. It is only necessary to determine the number of organisms which survive after suitable exposure of the treated culture to the drug to which it was originally sensitive. In this way a study has been made of the conditions which predispose an unencapsulated pneumococcal culture (R36A, derived from Type II[2]) to be transformed to high level streptomycin resistance.

An inoculum of 10^3–10^4 cells/ml. from a fresh 8-hour culture is grown in a neopeptone-meat infusion medium containing 0.2% of bovine plasma albumin and buffered at pH 7.5–7.8 with phosphate. After $3^1/_2$–4 hours at 37°C. the culture ordinarily responds to transforming factor following as little as 10 seconds exposure to 1–5 μg./ml. of DNA from a single high-step mutant strain resistant to over 2000 μg. streptomycin/ml. Exposure to DNA is terminated, usually after 5 minutes, by addition of *ca.* 1 μg./ml. of crystalline pancreatic desoxyribonuclease, a technique first employed by McCarty, Taylor, and Avery.[3] When 1 to 5 μg./ml. of DNA acts upon a young population of 10^6 to 10^7 cells/ml., conditions are sufficiently near saturation so that the number of susceptible cells is the limiting factor determining the number of transformants produced. The latter are scored by dilution into a nutrient broth containing 150 μg. streptomycin/ml. and a small amount of anti-R pneumococcus horse or rabbit serum globulin. The antibody causes the progeny of each viable unencapsulated cell to cluster as a single agglutinated colony which slowly settles during overnight growth and can be counted on the bottom of the culture tube. The demonstration of the validity of this bacterial counting method will be described in an article in preparation, in part by showing that correspondence with the more conventional methods is close. Within the appropriate range, 5 to 50 microcolonies per culture tube, inter-agglutination and overlapping of colonies are minor problems. During two to three years' use, the method has proven as reliable and accurate as agar plate counts within the same range, and it is more convenient and satisfactory for pneumococcus work. Scoring of transformants is done about

1–1½ hours after DNA action has been stopped, a time when they have already developed drug resistance but well before their number has begun to increase through multiplication.

After the original inoculum has grown for 3½–4 hours, a fraction of the population (0.2–5%, occasionally up to 17% in current practice) will respond to transforming factor. The fraction remains relatively constant during another 2 or more hours, then declines. Quantitative work has therefore confirmed similar qualitative conclusions derived from a study of the capsular transformation system.[3]

Once having been exposed briefly to a transforming agent, the susceptible cells are essentially irreversibly affected so that a variety of disturbances, changes in environment, inhibitors, etc. alter the ultimate number of transformants only to the degree that they modify growth of the whole culture. This stability permits the increments of transformable cells, which appear as the culture develops, to be estimated by comparing the yields after short and long exposures to DNA. The results of such studies tend to indicate that groups of susceptible cells are succes-

TABLE 1

NUMBER OF STREPTOMYCIN RESISTANT TRANSFORMANTS RECOVERED AFTER DIFFERENT SUCCESSIVE EXPOSURE TIMES

TIME OF EXPOSURE TO DNA, MIN.	YIELD OF S TRANSFORMANTS PER ML.[a]	ESTIMATED NUMBER OF TRANSFORMANTS DERIVED FROM AN ASSUMED SUSCEPTIBLE GROUP APPEARING AT TIME—					ESTIMATED NEW SUSCEPTIBLES DURING INTERVAL
		0–10 MIN., GROUP 1	10–20 MIN., GROUP 2	20–30 MIN., GROUP 3	30–40 MIN., GROUP 4	40–50 MIN., GROUP 5	
0–10	10,000	10,000	0	0	0	0	10,000
0–20	18,000	(10,000)	8,000	0	0	0	...
10–20	14,000	6,000	(8,000)	0	0	0	8,000
10–30	20,000	(6,000)	(8,000)	6,000	0	0	...
20–30	16,000	2,000	(8,000)	(6,000)	0	0	6,000
20–40	30,000	(2,000)	(8,000)	(6,000)	14,000	0	...
30–40	17,000	0	0	3,000	(14,000)	0	14,000
30–50	24,000	0	0	(3,000)	(14,000)	7,000	...
40–50	10,000	0	0	0	3,000	(7,000)	7,000

[a] Average of three determinations; estimated error ±10%. Values appearing in parentheses are taken from the experiment of the next preceding line.

sively developing and disappearing as the culture grows. This inference was first drawn from experiments of the type illustrated in table 1. Experiments indicate that it can be assumed that the DNA used is near a saturating concentration, and that cells in which transformation is once initiated will not revert. It now becomes possible to infer the number and subsequent history of these susceptible cells by observation based upon overlapping time intervals. It will be noted that there is evidence of new transformants whose development was initiated during the latter half of each 20-minute exposure. Since the total number of transformants remains relatively constant, it is concluded that some of those first susceptible later cease to be so. This mode of analysis of the data leads to the identification of the "groups" of cells indicated in the table. It may be seen that these groups in all cases seem to retain their susceptibility for somewhat less than 20 minutes; e.g., "group 2" appeared at some time between 10 and 20 minutes, and was no longer detected by 30 minutes. As a consequence of their being defined as made up of susceptible cells which "appear" during a certain time interval, these groups are of course

maximal at first and then diminish with time. It seems possible that other, physiologically more homogeneous, groups might be defined, if the appropriate physiological criteria were known, and that these groups would go through a rise and fall more symmetrically distributed in time. In any case, such classes derived from the present data and assumptions would almost certainly have the same periodicity as those indicated in table 1.

It should be added that longer periods of exposure to DNA give the expected greater proportion of transformants, provided that an effort is made to minimize and control the incidental effects of adding material and diluent upon the state of the culture. The conclusion from many experiments of the types mentioned is that groups or classes of organisms in the whole population are becoming transformable, remain so for about 15 minutes and then lose this potentiality. The proportion which the susceptible cells will reach in a given culture is reproducible under given conditions, but is influenced by changes in the medium and environ-

TABLE 2

INFLUENCE OF INHIBITING NUCLEIC ACID AND TIME OF EXPOSURE ON NUMBER OF TRANSFORMANTS RECOVERED

ADDITION OF INHIBITING DNA RELATIVE TO TRANSFORMING DNA	DURATION OF EXPOSURE TO		RECOVERED TRANSFORMANTS (NUMBER/ML.[a])	TRANSFORMATIONS INITIATED DURING EXPOSURE TO TRANSFORMING DNA		
	INHIBITING DNA	TRANSFORMING DNA		DURING 1ST MIN.	AFTER 1ST MIN.	PER CENT INHIBITION
no inhibitor	...	1 min.	44,000	44,000	...	...
no inhibitor	...	8	85,000	(44,000)	41,000	...
no inhibitor	...	15	84,000	(44,000)	40,000	...
no inhibitor	...	45	108,000	(44,000)	64,000	...
1 min. after	14 min.	15	54,000	(44,000)	10,000	75
simultaneously	15	15	20,000	10,000	(10,000)	77; (75)
6 min. before	7	1	250	250	...	100
6 min. before	14	8	13,000	(250)	13,000	68
6 min. before	21	15	15,000	(250)	15,000	63
6 min. before	51	45	19,000	(250)	19,000	70

[a] Average of three to six determinations; estimated error ±10%. Inhibiting calf thymus DNA, 100 μg./ml.; streptomycin transforming DNA, 1 μg./ml. Values in parentheses are taken from applicable experiment of a preceding line.

mental conditions in a manner which remains for the moment somewhat unpredictable. What seems of interest is that such relatively small fluctuations in the total number of transformants as are found in the (alternating) 10-minute experiments of table 1 seem to be attributable to rapidly changing populations of susceptible cells.

Further corroborative evidence has been obtained by taking advantage of the fact that a pneumococcus DNA not bearing streptomycin resistance factor, or indeed even calf thymus DNA in somewhat larger quantities, will compete with streptomycin resistance factor and interfere with transformation. As shown in table 2, inhibition by previously added thymus DNA is virtually complete for a short period (1 minute) after transforming DNA is added. Soon afterward, more cells become transformable, and from then on, the inhibition is only about 70%, corresponding to what is obtained when both DNA's are added together. This experiment was not carried out under conditions suitable for determining the rate

of transformation during prolonged exposure to DNA, since the cultures were not left undisturbed (see below). What this and other similar experiments do strongly indicate is that groups of susceptible cells (more or less uniform in responsiveness to either of two kinds of DNA) do appear from time to time within the growing culture.

The results just described confirm the impression that ability to react with DNA is a physiological property[3] and do not support the view that it is the result of a mutation[4] or a terminal change undergone by certain abnormal cells.[5] The strong suggestion of large groups of cells going through periodic change seemed to lead to the hypothesis that susceptibility might be connected with the bacterial division cycle, there being a certain phase of the cycle during which a limited fraction of the cells might be supposed to be most reactive with DNA. If this were true, any influence which might tend to produce even a modest degree of synchrony of cell division cycles in the whole bacterial population, would tend to increase the heights and depths of the "waves" of susceptibility. These should be relatively easy to demonstrate accurately enough, since the number of transformants should rise from zero to many thousand and fall back again, while the supposed regularities in the cell divisions would at best be expected to give only small periodic changes in the slope of the growth curve. For this reason, in most of our experiments, transformability has been the criterion for observation of possible synchronous behavior.

It was felt that the biochemical processes of cell growth and division would be disrupted in systematic fashion by temporary exposure of a growing culture to a temperature well below that at which these systems had achieved a steady state equilibrium. Certain of the enzyme systems should now be less able than others to transfer an amount of substrate equal to that supplied them, and there would then be a tendency for the metabolism to be selectively slowed or even temporarily halted at certain points. This should give opportunity for different cells to come into phase with each other, with regard to the slow processes leading to cell division.

Experimentation along this line showed that pneumococcal cultures could be rather easily influenced in this manner so as to give cyclical changes of transformability. A culture grown at 37°C., brought in 1 to 2 minutes to 25°C. and kept at this lower temperature for 15 minutes, then subdivided into aliquots and all returned to 37°C., is "phased" with respect to susceptibility to transformation. The number of susceptible cells present at the end of cooling is high, usually well above that of a replicate which had remained at 37°C. The data regularly are consistent with the supposition that cells initially transformable at the commencement of cooling remain so, and their number is usually materially augmented by cells newly become susceptible at the lower temperature. This state of the culture can be maintained at 25°C. for a period considerably longer than 15 minutes with little further gain, and with an eventual gradual diminution in the number of susceptible cells. In the investigation of a variety of times and temperatures of conditioning, the 15-minute period at temperatures of 22–25°C. has proved the most suitable.

After being returned to 37°C., however, the phased culture loses within 15–20 minutes virtually all its susceptibility to transformation, as indicated in figure 1. This change presumably represents the passage of the cells, accumulated in the susceptible state during cooling, into another state, characterized by low susceptibility. These effects of cooling and reincubation are reproducible and the drop in

susceptibility always occurs within 15 to 20 minutes after replacing the culture at the higher temperature. Such a fall may be attributed to the decay of almost any type of active state, superimposed upon another, rising, function.

There are signs, however, that transformability depends upon some type of cyclical process, since within a few minutes, susceptible cells begin to appear again and reach a new level, still higher in most instances than that previously attained. It would be difficult to ascribe the successive rises and falls in the curves of figure 1 to any combination of metabolic-accumulation-and-decay processes except one that is periodic in nature. As indicated, a second period of cooling may accentuate the apparent synchrony at subsequent peaks. This is the case when the cooling periods are separated by 40–45 minutes; when 15–30 minutes apart, the second low temperature treatment may obliterate or at least modify the effect of the first, giving a flat or highly irregular susceptibility-time curve.

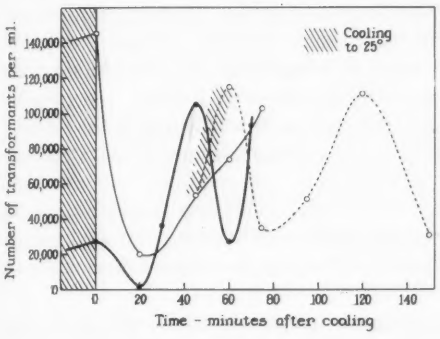


FIGURE 1

Cycles of transformability following conditioning of culture at 25 °C. for 15 minutes, then returning it to 37 °C. The heavy curve shows the result of one conditioning period. Another culture (lighter line) was given a similar treatment, and an aliquot of the same (dashed line) was submitted to a second conditioning 45 minutes after the end of the first one.

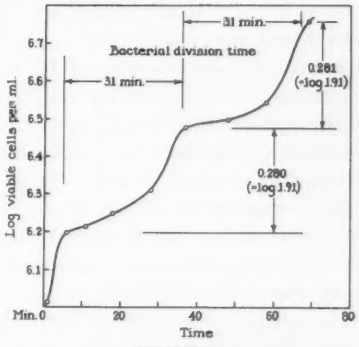


FIGURE 2

Course of pneumococcal growth immediately following 15 minutes' conditioning at 25 °C. Time shown is time of reincubation at 37 °C. Each point is the result of 4 to 6 determinations with estimated error of about ±10%. The indicated division time and logarithmic increments are formally derived from the drawn curve and only approximately determined by the experimental data.

All of these findings lead one to inquire to what extent these cyclical happenings may be related to the cellular division process. Under the conditions used here (complex media, late log phase, 37 °C.) the division time is from 30 to 35 minutes for pneumococci. The accumulation at 25 °C. of susceptible cells and their decay during the next 15 minutes at 37 °C. is not inconsistent with the time relations of cell division. The subsequent rise, however, is not easily to be correlated with the average division cycle of the population, since it appears to have a period of 40 to 60 minutes and be generally somewhat ill-defined and unpredictable. Nevertheless, the gradual flattening of later peaks farther removed from the initial temperature conditioning suggested that some type of partially synchronous cyclical

behavior had been brought about which was subsequently becoming more random.

It has become clear that nutritional influences, the presence of serum albumin, and the aeration of the culture incidental to redistribution of aliquots after conditioning, all are able to modify the rate of regain of susceptibility after it has been lost. These factors might obscure the supposed dependence of transformability upon a 30–35 minute division cycle. It has therefore seemed wise to return to the criterion of total cell growth for evidence of the synchronous metabolic cycles presumed to underlie the cycles of transformability.

Growth curves of temperature-conditioned pneumococcal cultures have given clear indications of partly synchronous discontinuities in cell division. The cultures were rapidly sampled at short time intervals into cold diluent broth, five to six replicate dilutions being made at each point. A curve showing parts of three waves of division is reproduced as figure 2. It will be seen that approximate two-fold increases in cell number were observed in two successive periods of about 30 minutes. Since untreated cultures give in general logarithmic doubling, spread uniformly throughout time intervals of the same order of magnitude, it is concluded that "phasing" of the culture has resulted in partially synchronous division. At present it still remains to be demonstrated whether the viable units separating at the steeper portions of the curve are individual cells, pairs of sister diplococci or, perhaps, fragments of short chains. Preliminary microscopic observations indicate that chain formation is not appreciable under the cultural conditions employed. It is not believed that simple errors of sampling and assay could destroy the significance of this demonstration of "waves" of division or of the similar observations made in other experiments of this type.

It will be noted that the first burst of division is a very rapid one appearing during the first few minutes after the culture is returned to 37°C. Later rises appear to be less "synchronous," but it should be pointed out that one-half of the total divisions are indicated as occurring within the last 5 minutes of each 30-minute cycle.

The correlation of these growth phases with transformability is imperfect, save perhaps for the phase immediately following reincubation, which is accompanied by a rapid loss of susceptibility by the transformable fraction of the population. The longer periods of the waves of susceptibility may be attributable to a different division rate in this part of the population, or to the additional influences affecting this property, as already mentioned. What does receive some support from the present observations is the hypothesis that bacterial division and transformability are both functions of cyclical metabolic activity which can be to some degree modified and rendered more nearly synchronous by environmental influence. If this can be verified, it may make possible other applications of bacteria in cytophysiological investigation.

Hegarty and Weeks[6] (and other authors quoted by them) have reported evidence of "fission waves" of bacteria in untreated cultures. It has recently been learned that Zeuthen[7] has independently also used temperature control as the basis for synchronizing division in the ciliate, Tetrahymena. Lastly, it should be mentioned that aeration and mixing, usually involved in the temperature conditioning as carried out heretofore, can in themselves tend to establish cyclical behavior in

pneumococci. The degree to which discontinuous oxygen shock can replace or modify the effect of temperature as a conditioning agent is under investigation.

Summary.—The transformability of pneumococcal cultures appears to be the result of cyclical processes occurring in individual groups in the population. These cycles, and the somewhat more regular ones of cell division to which they may be related, can both be accentuated and apparently rendered more synchronous, by the conditioning effect of appropriate environmental change.

[1] Hotchkiss, R. D., *Cold Spring Harbor Symp. Quant. Biol.*, **16**, 457–460 (1951).

[2] Avery, O. T., MacLeod, C. M., and McCarty, M., *J. Exp. Med.*, **79**, 137–157 (1944).

[3] McCarty, M., Taylor, H. E., and Avery, O. T., *Cold Spring Harbor Symp. Quant. Biol.*, **11**, 177–183 (1946).

[4] Alexander, H. E., and Leidy, G., *J. Exp. Med.*, **93**, 345–359 (1951).

[5] Alexander, H. E., discussion under Zinder, N., *Cold Spring Harbor Symp. Quant. Biol.*, **18**, in press.

[6] Hegarty, C. P., and Weeks, O. B., *J. Bact.* **39**, 475–484 (1940).

[7] Scherbaum, O., and Zeuthen, E., *Exp. Cell Research*, in press, 1954.

DOUBLE MARKER TRANSFORMATIONS AS EVIDENCE OF LINKED FACTORS IN DESOXYRIBONUCLEATE TRANSFORMING AGENTS

By Rollin D. Hotchkiss and J. Marmur*

LABORATORY OF PATHOLOGY AND MICROBIOLOGY, THE ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH, NEW YORK

Communicated by R. J. Dubos, December 23, 1953

The ability of certain bacterial desoxyribonucleic acid (DNA) preparations to transfer heritable characters of a donor strain over to a related strain of bacteria[1] has suggested the implication that chromosomal DNA may play the major role in genetic determination in the higher organisms. It has not been clear, however, whether the "transforming agents" should be considered the analogs of the genes, of whole chromosomes, or of some other level of organization. That they do not convey the total genetic complement of a donor strain to a recipient cell was already suggested by Langvad-Nielsen, who showed with the early techniques of in vivo transformation that capsular transformants of pneumococcus retained the same sulfonamide resistance as the recipient strain, acquiring, so far as was known, only the capsular property from the donor strain.[2] It has more recently been shown that drug resistance can indeed be transferred through DNA from a donor strain, but that the resistance factor and the capsular property are not in general both demonstrable in one and the same transformed cell. This is true both for streptomycin resistance and several agents conferring different quantitative levels of penicillin resistance in *Diplococcus pneumoniae*.[3] It was concluded that individual DNA particles had carried individual unit characters into the recipient cell. These results and others indicated close correspondence between the unit tranforming principle and the bacterial gene.[4]

It was estimated that the rare doubly-transformed (encapsulated and penicillin resistant) cell appeared in a proportion approximately in the range of the product of the frequencies of the two individual transformations within the population.

TABLE 1

RECOVERY OF MULTIPLE TRANSFORMANTS FROM PNEUMOCOCCI TREATED WITH VARIOUS TRANSFORMING AGENTS

EXPT.	EXPOSURE TO AGENT		OBSERVED TRANSFORMATION TO: SINGLE PROPERTIES, % OF POPULATION		MULTIPLE PROPERTIES, %	EXPECTED DOUBLE TRANSFORMATIONS: ROUTE*	PERCENTAGE EXPECTED	FOUND EXPECTED
1	DNA$_{MS}$	5 min.	S: 2.0	M: 0.32	MS: 0.091	M × S	0.0061	14
	DNA$_{MS}$	10 min.	4.3	0.40	0.17	M × S	0.017	10
	DNA$_{MS}$	15 min.	5.4	0.67	0.16	M × S	0.036	4.4
2	DNA$_{MS}$	30 min.	S: 2.6	M: 2.6	MS: 0.40	M × S	0.067	6
3	DNA$_{M}$ + DNA$_{S}$	5 min.	S: 1.4	M: 0.87	MS: 0.003	M × S	0.012	0.2
		10 min.	S: 2.8	M: 0.83	MS: 0.0066	M × S	0.023	0.29
4*a*	DNA$_{PSMP}$	10 min.	S: 4.0	M: 2.6	MS: 0.41	M × S	0.10	4.1
			F: 4.0		SF: 0.22	S × F	0.16	1.4
					MF: 0.0058	M × F	0.10	0.06
					MSF: 0.0071	M × S × F	0.0041	1.7
						SF × M	0.0057	1.2
						MS × F	0.016	0.44
						MF × S	0.00023	31
4*b*	DNA$_{PSMP}$	10 min.	S: 2.7	M: 1.6	MS: 0.10	M × S	0.043	2.3
			P: 0.67		SP: 0.0071	S × P	0.018	0.4

This correspondence indicated that the double transformations could have resulted simply from successive transformation events brought about by two independent single factor agents. Thus, if approximately 1% of the cell population responds to either agent, one may expect 1% of 1%, that is, 0.01% or 1 per 10,000 to have had the opportunity to react with both.

The development in this laboratory of procedures for dealing with several new transformable characters of *Diplococcus pneumoniae* has permitted a more adequate search for pairs of characters which might not behave independently in a transformation process. The characters used in this survey have included penicillin resistance (first step, "P"), streptomycin resistance (high step, "S"), and two not previously described; sulfanilamide resistance ("F") and mannitol utilization ("M", the ability to grow and metabolize mannitol in media in which this sugar alcohol is the limiting energy source). Cells bearing this last property are quantitatively determined in a defined medium in which a mixture of pure amino acids[5] approximating the composition of casein hydrolysate with a small quantity (0.005%) of peptone (Tryptone, Difco) dialysate are supplemented by the vitamin and mineral mixtures described by Adams and Roe.[6] With the addition of mannitol (0.3%) and sodium pyruvate (about 0.01%) this medium fosters selective growth of mannitol mutants or transformants. The drug resistant strains are scored and isolated in neopeptone-meat infusion broth made selective by addition

of either 0.05 unit/ml. crystalline benzylpenicillin, or 150 μg./ml. streptomycin, or in casein hydrolysate peptone media containing 15 μg./ml. of sulfanilamide. Combinations of the drugs in these media allowed scoring for two properties. To all of the media were also added small amounts of anti-pneumococcal antibody globulin, which serve to isolate the growth in discrete colonies. The counting of colonies arising from appropriate dilutions in such media is the basis for the scoring method to which reference has already been made.[7]

It is found that two properties, mannitol utilization, M, and streptomycin resistance, S, are frequently both introduced into the same transformed cells during a short exposure to DNA preparations (1–5 μg./ml.) from donor strains bearing both traits. As illustrated in table 1, experiments 1 and 2, the number of MS transformants is commonly as great as one-fourth the number of M cells produced, although it is usually smaller in proportion to the S transformants. In these experiments, the number found is from 5 to 15 times the number of double transformants that would be "expected" on the assumption that they arose by successive or simultaneous random (S × M) transformations. These findings are similar to those from many other experiments and are illustrated again in the multiple factor transformations of 4*a* and 4*b*, table 1. Here, values in the last column near to, or less than, unity indicate those routes by which (in any sequence) the multiple transformants could with reasonable probability have arisen. It will be seen that the assumption of random independent transformations is satisfactory for most pairs of characters except M and S; even in the case of the rare triple transformant, MSF, the most acceptable hypothesis is that M and S were somehow acquired nonrandomly, and the resultant MS cells were independently transformed with respect to the F property.

The M and S characters have no apparent obligatory connection in the metabolism (phenome) of pneumococci, each property having appeared independently of the other a number of times in spontaneous or induced mutations, or through transformation. Therefore, it becomes appropriate to ask whether the nonrandomness of the MS transformations is due to non-randomness in the responsiveness to the two transforming factors within susceptible bacterial populations, or rather to some structural relationship linking the two DNA transforming factors. A mixture of transforming factors, DNA_M and DNA_S, from separate M and S strains, should not be able to induce double transformations to MS if the latter principle is operating. Experiment 3 of table 1 illustrates that mixed DNA's gave only a few MS transformants, well below the number that independent transformations could have provided.

These calculations disregard the considerations indicated elsewhere[7] that some of the cells in a culture are more predisposed to transformation than others, but they are probably not seriously unreliable as long as those cells potentially capable of responding, biologically, to a transforming agent are a relatively large fraction of the population. Actually the ability of one DNA factor to prevent the utilization of another by the same cell,[7] seems to be the cause of larger deviations, in the opposite direction. Multiple transformations believed to be unlinked occur less frequently than random, as indicated by factors in Table 1 varying from 0.06 to 0.44.

The results therefore indicated strongly that DNA derived from an MS strain

confers the M and S transforming agents together in some kind of association not found in a mixture of M and S transforming agents derived from separate cells. The possibility was considered that the association might be, e.g., an aggregation which the separate agents underwent during reprecipitation, etc., as the DNA preparation is made. Accordingly, a mixed culture of singly marked M and S cells (ratio M/S = 1.3) was grown, and DNA prepared from it by the usual procedure. This preparation behaved like the artificial mixture of separately isolated M and S DNA's, giving virtually no MS transformations (e.g., when recovery of *ca.* 100 cells/ml. would have indicated non-random transformation, none were recovered).

It appears, therefore, that the special relationship of the S to the M transforming factor (but not to certain others) is the result of a special relation existing within the donor MS cell. It is natural to suppose that the association may be the equivalent of genetic linkage. The incompleteness of the linkage would have an obvious parallel in "crossing-over" which separates, to varying degrees, genetically linked factors in the higher forms. It might also be due to heterogeneity among the DNA particles derived from the MS strain. The manifestation of such phenomena by a preparation of DNA raises profound questions which cannot be answered as yet, concerning the relation of these properties of DNA to those of the chromosomal linkage groups in higher species. In particular, if one now is inclined to ask whether chromosomes also owe their linkage patterns to DNA structures rather than to connected bands of nucleoproteins, one is correspondingly obliged to question whether bacterial DNA's showing linked properties can be prepared as free of protein as the capsular transforming preparation, which contained none, or at most less than 0.02%.[8] There is no indication that they do contain protein; this question will be further investigated. Linked transformations have, in any case, been brought about by DNA preparations resistant to the autolytic proteolysis and alcohol precipitation, and sustaining the removal of protein by surface denaturation, which occur in the usual procedure of purification. Furthermore, all of their transforming activities are lost in brief treatment with crystalline desoxyribonuclease.†

It is found that the same type and degree of linkage of M and S factors is exhibited by DNA from two MS strains, one of which had acquired its S property through a discrete spontaneous mutation (expt. 2, table 1) and the other of which was an M strain transformed by DNA from an S strain (expt. 1). Thus, the site taken up by the S determinant within the cell bears the same relation to the M factor when it arrives by mutation and when it is introduced by tranformation. This fact appears to give substance to the concept of a unique genetic "locus" in bacteria. The locus of the S factor may be supposed to be within the same DNA particle which bears the M factor.

Further support, for the picture outlined above, is given by experiments on the transformations to drug sensitivity. Drug-sensitive bacteria, arising from a resistant population in small numbers, cannot satisfactorily be detected by a differential survival count. However, if streptomycin-sensitive organisms were produced in large proportion among the cells simultaneously acquiring the mannitol-utilization trait, they could easily be identified by examining an appropriate number of single-colony strains of the M transformants selected from the untransformed population by virtue of their M property. As indicated in table 2, the action of DNA from an

ordinary M (streptomycin-sensitive, "S−") strain sometimes confers, along with the M property, streptomycin sensitivity, displacing or inactivating the S determinant of resistance already present. Analogous results are recorded for the action of S factor in displacing M from an M strain. These findings again demonstrate that there is a unique locus within the bacterial cell capable of determining metabolic functions leading either to streptomycin resistance or to sensitivity. They also re-emphasize that this site and a corresponding one determining mannitol utilization stand in special relationship to each other within or upon the cell DNA particles.

Finally, it may be mentioned that another pair of pneumococcal characters, S and F, seem to a lesser extent also to have given linked transformations, as indicated both in table 2 and in experiment 4, table 1.

TABLE 2

LINKED TRANSFORMATIONS INVOLVING LOSS OF A SELECTIVE MUTANT PROPERTY

CHARACTERS PRESENT			TRANSFORMANTS	SINGLE COLONY STRAINS TESTED	
DONOR DNA	RECIPIENT CELLS	EXPT. NO.	SELECTED FOR	NUMBER OF STRAINS	NUMBER OF DOUBLE TRANSFORMANTS FOUND
M(S−)	S(M−)	1	M	19	4 M(S−)
		2	M	9	1 M(S−)
		3	M	7	none
		4	M	10	1 M(S−)
		5	M	4	1 M(S−)
S(M−)	M(S−)	6	S	18	7 S(M−)
		7	S	24	1 S(M−)
		8	S	30	4 S(M−)
MP(S−)	M(S−)	9	P	10	(none M−)
S(M−)	MP	10	S	4	2 SP(M−)
M(F−P−)	PF(M−)	11	M	3	1 PM(F−)
S(M−F−P−)	PFM(S−)	12	S	4	1 PMS(F−)

Summary.—Several pairs of transforming agents give randomly distributed independent transformations in pneumococci. In contrast, desoxyribonucleate derived from a strain bearing the two mutant properties, mannitol utilization and streptomycin resistance, gives a high proportion of double transformations to both of these selective properties. The results suggest the material existence in the bacteria of specific genetic loci, involving these determinants in some mode of linked association within the desoxyribonucleate particles of the doubly marked strains. These sites may also carry factors associated with streptomycin sensitivity or incapacity to utilize mannitol, similarly linked, and capable of displacing the selective mutant property from a cell already bearing it.

* Public Health Research Fellow, 1952–53.

† Of no little interest in itself is the fact that mannitol utilization, an adaptive system in the M strains of pneumococcus, is conferred equally well by DNA of these strains, whether they are in the adapted or the unadapted state.

[1] Avery, O. T., MacLeod, C. M., and McCarty, M., *J. Exp. Med.*, **79**, 137–157 (1944).

[2] Langvad-Nielsen, A., *Acta Path. Microbiol. Scand.*, **21**, 362–369 (1944).

[3] Hotchkiss, R. D., *Cold Spring Harbor Symp. Quant. Biol.*, **16**, 457–460 (1951).

[4] Hotchkiss, R. D., *Exp. Cell Research*, suppl. **2**, 383–390 (1952).

[5] Rockenmacher, M., James, H. A., and Elberg, S. S., *J. Bact.*, **63**, 785–794 (1952).

[6] Adams, M. H., and Roe, A. S., *Ibid.*, **49**, 401–409 (1945).

[7] Hotchkiss, R. D., these PROCEEDINGS, **40**, 49–55 (1954).

[8] Hotchkiss, R. D., in *Phosphorus Metabolism*, edited by W. D. McElroy and Bentley Glass, Johns Hopkins Press, Baltimore, Vol. 2, 426–436 (1952).

---

*THE ISOLATION AND IDENTIFICATION OF THE MAJOR ANION FRACTION OF THE AXOPLASM OF SQUID GIANT NERVE FIBERS**

BY BERNARD A. KOECHLIN

BIOLOGY DEPARTMENT, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated by Francis O. Schmitt, December 21, 1953

It has been known for some time that a large fraction of the anions in the axoplasm of squid giant nerve fibers and in lobster nerves are organic in nature.[1–5] It could be deduced that the organic ions have an equivalent weight of about 100.[2] It was also demonstrated that lobster nerve contains high concentrations of free amino acids, particularly aspartic acid and alanine.[5] However, even after this discovery the anion deficit remained considerable; some 60 per cent of the anions remained to be identified. From analyses of the free acidic amino acid content of invertebrate nerves, particularly those of Carcinus, Lewis[6] believed to have demonstrated acid-base balance to within about 10 per cent. These analyses were made on intact nerves rather than on extruded axoplasm. Therefore, acid-base distribution between intracellular and extracellular phases could only be inferred. The data on electrolyte balance were thoroughly reviewed by Hodgkin.[7]

Because of the possible significance of the axonic anions in the biochemical processes and bioelectric properties of the nerve fiber, efforts were made to determine the identity of the missing organic anion. It is believed that this has now been accomplished in the case of squid axoplasm. In this paper are reported the chief results of the analyses and identifications; the details will be reported subsequently.

Method of Isolation of the Anion.—Axoplasm from squid giant fibers was prepared in the manner described by Maxfield.[8] Individual samples consisted of aqueous suspensions of the pooled axoplasm from the nerves of 20 to 60 squid. The amount of axoplasm in each preparation was determined from dry weights taken for each pool and from our corrected value for the water content of extruded axoplasm (86.5 per cent). For the isolation of the free anions either the aqueous suspension of axoplasm was dialyzed against distilled water at 0°C. or precipitated and extracted with 60 per cent ethanol at about −10°C. Both methods led to satisfactory extraction and essentially identical results. Isolation of the unidentified anion in practically quantitative yields was achieved by subjecting the extract to a sequence of ion exchange column operations as follows: passage through Dowex-50, partly in its silver form, removed cations and dipolar molecules and retained the insoluble silver salts of chloride and phosphates; passage through Amberlite IR-45† then removed the anions, allowing uncharged molecules to pass through; after elution of the anions from the Amberlite with base, the salts thus formed were passed through Dowex-50, partly in its mercury form, retaining the dicarboxylic acids and yielding the free remaining acid. In a typical isolation about 15 to 20 mg. of this acid, in

practically pure form, were obtained from the axoplasm of the nerves of about 100 squid.

Identification of the Anion.—The hygroscopic free acid which crystallized with considerable difficulty has an equivalent weight of 135 ± 10, as determined on the barium salt. No measurable optical activity was observed. Titration curves indicated a strong acidic group with a pK close to 2.0; no secondary inflection at higher pH, corresponding to a multivalent acid, was observed. That the acid is not carboxylic was suggested by the fact that no ketone, ester, or lactone group, or unsaturation could be demonstrated, such as might confer strong acidity upon a carboxylic acid.

The clue to the identification of the anion came with the demonstration of a sulfonate group. The axoplasmic anion, on digestion with nitric acid, consistently liberated one mole of sulfuric acid per equivalent of acid. Since this did not occur

TABLE 1

ACID-BASE BALANCE IN SQUID NERVE AXOPLASM

SUBSTANCE	CONCENTRATION μ EQUIV. PER G. OF AXOPLASM	ANALYTICAL METHOD USED
Anions		
Chloride	150 ± 20	Conway[9]
Phosphates	35 ± 3	Lowry and Lopez[10]
Aspartic acid	65 ± 3	....
Glutamic acid	10 ± 2	Moore and Stein[11]
Fumaric acid	20 ± 5	Martins and Nitz-Litzow[12]
Succinic acid		
Isethionic acid[a]	230 ± 20	Cuthbertson and Tomsett[13]
Total anions	500 ± 20	....
Total base	520 ± 20	Ion exchange titration[14]

[a] Sulfate liberated from anionic sulfonate.[13]

as a result of acid or alkaline hydrolysis the material could not be a sulfate ester. The presence of one primary hydroxyl group attached to a second carbon atom was indicated from the results on oxidation with chromic acid: consistently, two equivalents of oxygen were consumed per equivalent of anion under formation of divalent carboxylic sulfonic acid demonstrated titrimetrically. No ketone, aldehyde, or sulfuric acid resulted. The presence of sulfonate and hydroxyl groups was confirmed by infra-red spectrographic analysis.

These facts, together with the equivalent weight figure, strongly suggested that the substance is 2-hydroxyethane sulfonic acid (or perhaps its next higher analog). Evidence confirming this conclusion was obtained by demonstrating, by infra-red spectrography, the identity of the anion with a sample of 2-hydroxyethane sulfonic acid (isethionic acid) prepared by deamination of taurine. Both samples were prepared as the methyl esters. All the bands observed in the spectrum of the anion sample were present also in the spectrum of the synthetic compound.

Acid-Base Balance.—From table 1 it will be seen that the amount of isethionic acid present in axoplasm is just sufficient, when added to the other anions demonstrated to be present, to balance the base. Detailed documentation of these experiments will be provided elsewhere.

Experiments designed to determine whether the predominant anion of lobster and other nerves is also isethionic acid are in progress as well as investigations of the possible role of this substance in the biochemical and physiological processes in nerve.

Acknowledgments.—The author is indebted to Professor F. O. Schmitt for critical interest and support of this work, to Dr. J. F. Scott for the infra-red analyses, and to Dr. M. Maxfield and the dissection team for help in providing squid axoplasm.

* These studies were aided by a contract between the Office of Naval Research, Department of the Navy, and the Massachusetts Institute of Technology, NR119-100; a research grant, B-24, by the National Institute of Neurological Diseases and Blindness of the National Institute of Health, U. S. Public Health Service; and a grant from the Trustees under the wills of Charles A. King and Marjorie King.

† Kindly supplied by Dr. R. Kunin, Rohm and Haas Company.

[1] Fenn, W. O., Cobb, D. M., Heguaner, A. H., and Marsh, B. S., *Am. J. Physiol.*, **110**, 74 (1934).
[2] Bear, R. S., and Schmitt, F. O., *J. Cell. Comp. Physiol.*, **14**, 205 (1939).
[3] Schmitt, F. O., Bear, R. S., and Silber, R. H., *Ibid.*, **14**, 351 (1939).
[4] Silber, R. H., and Schmitt, F. O., *Ibid.*, **16**, 247 (1940).
[5] Silber, R. H., *Ibid.*, **18**, 21 (1941).
[6] Lewis, P. R., *Biochem. J.*, **52**, 330 (1952).
[7] Hodgkin, A. L., *Biol. Revs.*, **26**, 339 (1951).
[8] Maxfield, M., *J. Gen. Physiol.*, **37**, 201 (1953).
[9] Conway, F. J., *Biochem. J.*, **29**, 2221 (1935).
[10] Lowry, O. H., and Lopez, J. A., *J. Biol. Chem.*, **102**, 421 (1946).
[11] Moore, S., and Stein, W. H., *Ibid.*, **192**, 663 (1951).
[12] Martins, C., and Nitz-Litzow, D., *Chem. Ber.*, **85**, 605 (1952).
[13] Cuthbertson, D., and Tomsett, S., *Biochem. J.*, **25**, 1237 (1931).
[14] Koechlin, B. A., in preparation.

---

*DIELECTRIC INCREMENTS IN AQUEOUS SOLUTIONS OF SYNTHETIC POLYELECTROLYTES**

BY HOWARD M. DINTZIS, J. L. ONCLEY, AND RAYMOND M. FUOSS

UNIVERSITY LABORATORY OF PHYSICAL CHEMISTRY RELATED TO MEDICINE AND PUBLIC HEALTH, HARVARD UNIVERSITY; AND STERLING CHEMISTRY LABORATORY, YALE UNIVERSITY

Presented before the Academy, April 28, 1953; Communicated December 23, 1953

Dielectric properties in aqueous solution have been reported for two general classes of charged macromolecules. The first of these classes, that of the globular proteins, has been studied in some detail; the experimental results have been consistent with reasonable assumptions concerning the size and shape of the molecules and the distribution of ionized groups.[1, 2] Measurements on the second class, that of extended polyelectrolytes, have been limited until very recently to the nucleic acids and similar complex biological materials of relatively unknown structure.[3, 4, 5] The electrical behavior of solutions of these two classes of macromolecules exhibits striking differences.

Some conclusions concerning the size and shape of the nucleic acids have been deduced from dielectric measurements, using a model which represents nucleic acid

molecules as rigid rotating dipoles, although there is little independent evidence supporting such an assumption. It is therefore of interest to study the dielectric properties of polyelectrolytes of known molecular structure and charge distribution in an effort to correlate macroscopic and molecular parameters. Studies of this type have been reported recently for a synthetic polyelectrolyte[6] and for carboxymethylcellulose.[7]

The poly-4-vinyl-N-*n*-butylpyridinium bromide used in our work was prepared from the middle fraction of a sample of poly-4-vinylpyridine whose reduced viscosity (η_{sp}/C) at 0.200 g./100 ml. in 95% ethanol was 6.18. We estimate from this

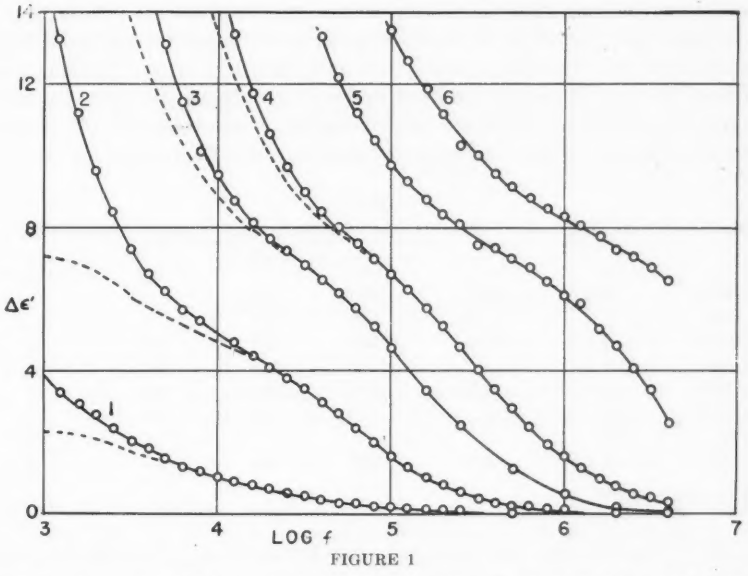


FIGURE 1

Dependence of dielectric constant of aqueous solutions of poly-4-vinyl-N-*n*-butylpyridinium bromide on concentration and frequency: 1, $g = 0.0010$; 2, 0.0049; 3, 0.0162; 4, 0.050; 5, 0.40, and 6, 2.40 g./l.

an approximate molecular weight of 2×10^6. Quaternization was effected by treating the polymer with *n*-butyl bromide in tetramethylenesulfone; the product contained 25.9% Br, which corresponds to 60.7% quaternization. The salt was slightly acid, potentiometric titration indicating that approximately 1% of the nitrogen was quaternized by hydrogen bromide. The salt was dried by lyophilization from water, followed by storage in a desiccator over solid potassium hydroxide. A stock solution was prepared by dissolving a weighed quantity of material in conductivity water made by resin deionization of distilled water. Dilute solutions were prepared by weight dilution of the stock solution. For the most dilute solutions, all operations were carried out in the absence of carbon dioxide in order to eliminate the effects of extraneous electrolyte.

Electrical measurements were made by the bridge method over the frequency range from 1 kilocycle to 4 megacycles. The type of apparatus used has been

described previously.[8, 9] In order to minimize polarization and other errors resulting from the wide frequency and conductivity range involved, calculations were based on the difference in behavior between the polyelectrolyte solution and a reference solution of potassium bromide of approximately the same conductance. Electrode polarization corrections were made at low frequencies by the empirical method of Oncley.[10]

The dielectric increment $\Delta\epsilon'$ was measured over the concentration range, 0.0010–2.40 grams per liter; concentrations in these units are designated by the symbol g. The results are summarized in figure 1, where $\Delta\epsilon'$ is plotted against the logarithm of frequency for the six concentrations measured. The solid curves represent the original observations, while the dotted curves are corrected for polarization. It will be seen that the effect of electrode polarization capacity was small for frequencies above 10^5 cycles per second. The polarization, however, becomes a large fraction of the total capacity increment at frequencies below 10^4 cycles per second; it is therefore necessary to regard the increments in the region of 10^3 cycles per second as tentative values on account of uncertainties in the correction.

TABLE 1

DEPENDENCE OF DIELECTRIC INCREMENT ON FREQUENCY AND CONCENTRATION AT 25° C.

		$\Delta\epsilon'/g$				
g	$10^6\kappa$	$f = 10^3$	10^4	10^5	10^6	4×10^6
0.0010	0.6	2200	1000	100	...	...
0.0049	1.4	1400	980	310	...	...
0.0162	3.4	1050	560	270	20	6
0.0480	11.3	...	...	140	37	6
0.0500	8.9	820	280	135	32	6
0.400	64	...	...	25	15	6
2.40	323	...	...	5.6	3.5	2.7
0.048[a]	5.6	290	200	100	15	3

[a] Measured at 0.0°C.

Values of the dielectric increment per gram per liter, $\Delta\epsilon'/g$, are given in table 1 for various frequencies and concentrations; conductances at one kilocycle are given (ohm^{-1}cm.$^{-1}$) under the heading κ in the second column. It will be noted immediately that the dielectric increment per gram for the polybromide covers an enormous range of values. The electrical behavior of the synthetic polyelectrolyte stands in marked contrast to that of amino-acids, peptides, and proteins in two distinct respects. First, for the latter group of naturally occurring substances,[1] $\Delta\epsilon'/g$ has a maximum value of about 2, whereas the synthetic polyelectrolyte solutions show a value of $\Delta\epsilon'/g$ of several thousand at low frequencies and concentrations. Secondly, the quantity $\Delta\epsilon'/g$ is almost independent of concentration in the case of amino-acids, peptides, and proteins, whereas in the case of the polyelectrolyte, a large concentration dependence is evident from the data in table 1. Natural polyelectrolytes have been reported to have relatively large values of $\Delta\epsilon'/g$, some typical values being 200 for sodium thymonucleate,[4] 80 for thymonucleohistone,[5] and 120 for sodium carboxymethylcellulose.[7] These materials also show a strong dependence of $\Delta\epsilon'/g$ upon concentration at constant frequency.

It was possible to measure the changes in conductance ($\Delta\kappa$) with frequency for most of the solutions. The values for the more concentrated solutions are shown in

figure 2. This type of behavior is consistent with a purely a.-c. dielectric loss in the solution. The relationship to changes in dielectric increment can best be seen in the Cole plots of figure 3, where values of $\Delta\epsilon''$ are calculated from the conductance

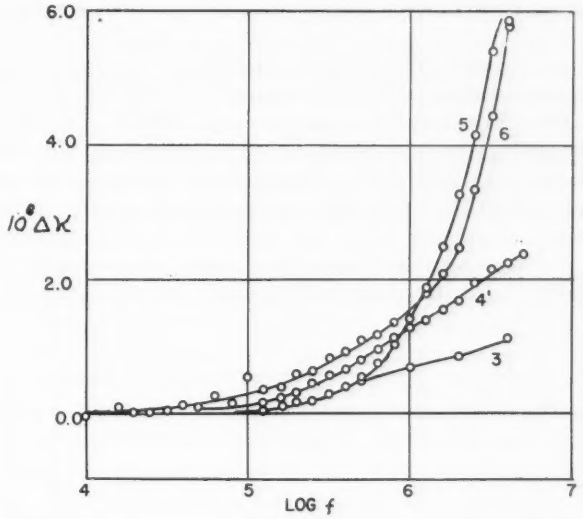


FIGURE 2

Increment of conductance of polyelectrolyte solutions as a function of concentration and frequency: 3, $g = 0.0162$; 4', 0.048; 5, 0.40; 6, 2.40 g./l.

increment from the equation $\Delta\epsilon'' = 1.80 \times 10^{12}\ \Delta\kappa/f$. The curves for $g = 0.048$ and 0.40 show that the in-phase component of the complex dielectric constant, $\Delta\epsilon''$, is due to the same mechanism as that which produces the dielectric increment $\Delta\epsilon'$. These two curves approximate a circular arc with center below the $\Delta\epsilon'$-axis; in other

TABLE 2

EFFECT OF ADDED POTASSIUM BROMIDE AT 25°C.

$10^3 g$	10^3 [Br]	r	$\Delta\epsilon'$		
			$f = 10^3$	10^4	10^5
4.90	1.59	0.0	7	4.8	1.6
4.75	1.54	0.8	6	3.0	1.1
4.60	1.49	1.6	5	2.2	0.8
4.48	1.45	2.4	5	2.3	0.7
4.35	1.41	10.0	3	1.1	0.5

words, the maximum in $\Delta\epsilon''$ is less than half the estimated low frequency limit of $\Delta\epsilon'$. We therefore conclude that a distribution of relaxation times is necessary to describe the dielectric dispersion. The sharp rise toward lower frequencies for the curve at $g = 2.40$ is possibly due to polarization, for which no correction was made in this plot. The high frequency end of this curve, however, appears to be approaching normal behavior.

Conductance measurements at high field strength[11] on solutions of another sample of this polysalt show a dispersion of the Wien effect with a time constant of the order of microseconds. This observation is independent confirmation of the existence of a dispersion of dielectric constant and conductance with frequency, and also of the order of magnitude of the relaxation time deduced from the inflexion points of figure 1.

The addition of simple salt (potassium bromide) to the polyelectrolyte solution produced a marked decrease of dielectric increment. The data are given in table 2, where g denotes polyelectrolyte concentration in g./1000 ml., [Br] bromide ion concentration originally present as gegen ion in the polyelectrolyte in equivalents per liter, and r is the ratio [KBr]/[Br]. A similar effect has been observed for sodium thymonucleate solutions.[4] This behavior is in sharp contrast to that of amino-acids, peptides, and proteins, which are little influenced by the presence of electrolyte, but is analogous to the suppression of other polyelectrolyte characteristics (such as viscosity[12] and streaming birefringence[13]) by the addition of simple electrolyte.

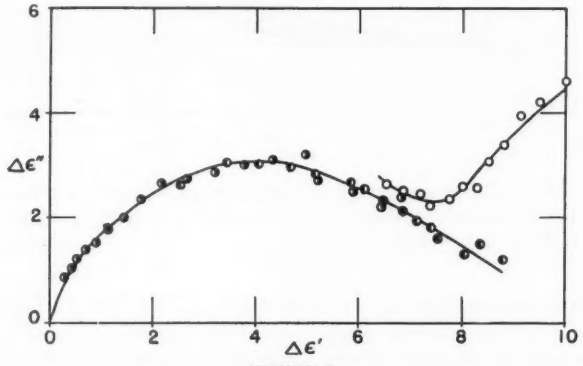


FIGURE 3

Cole plots from data of figs. 1 and 2: open circles, $g = 2.40$; left black, $g = 0.40$; right black, $g = 0.048$.

The effect of temperature upon the dielectric increment and conductance was measured at one concentration, $g = 0.048$. Measurement at 0°C. and 25°C. (table 1) showed that the pattern of the frequency dependence of dielectric increment was the same at both temperatures but the dispersion curve was shifted by a factor of 0.40 toward lower frequencies at the lower temperature. At least part of this shift can be ascribed to viscosity changes, but no data for comparison are available on the viscosity of polyelectrolytes as a function of temperature. The Cole plot for the data at 0°C. coincided with that for the 25°C. data.

Since our data do not permit evaluation of the low frequency limit for $\Delta\epsilon'$, it was not possible to obtain accurate relaxation times. It may be seen from figure 1, however, that the region of dielectric dispersion covers too wide a frequency range to be explained in terms of a single relaxation time. The dielectric dispersion shifted toward higher frequencies as the concentration was increased, in agreement

with the behavior of the natural polyelectrolytes, but directly opposite to the behavior of the proteins.

The similarity in all respects between the dielectric behavior of our synthetic polyelectrolyte and that of the natural linear polyelectrolytes suggests that the same molecular mechanisms are responsible for most of the dielectric behavior. Because of the flexibility of the molecular structure of the linear polymer and the purely cationic nature of poly-4-vinyl-N-*n*-butylpyridinium bromide, it seems extremely unlikely that there could exist a permanent dipole moment sufficiently large to account for the observed effects. A fluctuation mechanism of the specific type proposed by Kirkwood and Shumaker for proteins, which postulates a non-vanishing mean square moment due to the exchange of protons among equivalent basic sites, would not seem to apply here because the charges are fixed to the pyridinium nitrogens.

The polyelectrolyte molecule may be regarded as a flexible, highly charged thread of great length surrounded by a relatively diffuse cloud of bromide ions, the spatial distribution of bromide ions being determined by the electrostatic potential. At very low concentrations the molecules will tend to stretch out and assume a rod-like shape because of intramolecular electrostatic repulsion between neighboring segments of the chain. On account of the electrostatic attraction of bromide ions to the chain, the latter will tend to cluster near it, resulting in high concentration of mobile charge close to the macromolecule and low concentration in regions relatively far away. A detailed analysis of the potential of an infinite rod-like molecule and the distribution of the counter ions has been presented.[14] If the electrostatic screening due to small ions is increased by increasing the polymer concentration or by the addition of simple salts, the intersegmental repulsion along the chain will decrease and Brownian motion will lead to the contraction of the molecule into a random coil.

In view of these considerations, the dilute aqueous solution of polyelectrolyte may be regarded as a heterogeneous medium, composed of regions of relatively high conductivity in the vicinity of the macromolecule, suspended in a continuous medium of considerably lower conductivity. The size and shape of the regions of high conductivity and the dependence of conductivity upon distance will be related in a very complicated manner to factors such as concentration, chemical constitution, molecular weight, charge density, and the presence of simple salts. We shall now show that such a model can lead to a dielectric dispersion which agrees in both magnitude and frequency with that observed experimentally. In addition, the dependence of electrical behavior on the concentration of polymer and on that of added simple salts may be qualitatively interpreted.

The electrical properties of a polyphase dielectric consisting of a continuous medium of dielectric constant ϵ_1' and conductivity σ_1 in which are suspended small spheres of a second phase of dielectric constant ϵ_2' and conductivity σ_2 was solved by Maxwell [15] and generalized by Wagner[16] (here, σ denotes conductivity in e. s. u.). It was shown[17] that if the suspended spheres occupy a volume fraction, $p \ll 1$, the bulk dielectric constant is given by the equation

$$\epsilon = \epsilon_1[2\epsilon_1 + \epsilon_2 - 2p(\epsilon_1 - \epsilon_2)]/[2\epsilon_1 + \epsilon_2 + p(\epsilon_1 - \epsilon_2)].$$

Here, both ϵ_1 and ϵ_2 are complex quantities of the form $\epsilon_1 = \epsilon_1' - j\epsilon_1''$ and $\epsilon_2 =$

$\epsilon_2' - j\epsilon_2''$. This relationship can be put into the following normal form by separating real and imaginary components:

$$\epsilon = \epsilon_\infty \left(1 + \frac{k}{1 + \omega^2\tau^2} - \frac{jk\omega\tau}{1 + \omega^2\tau^2}\right), \tag{1}$$

where

$$\epsilon_\infty = \epsilon_1' \left(1 - \frac{3p(\epsilon_1' - \epsilon_2')}{2\epsilon_1' + \epsilon_2'}\right)$$

$$k = 9p(\sigma_1\epsilon_2' - \sigma_2\epsilon_1')^2/\epsilon_1'(2\epsilon_1' + \epsilon_2')(2\sigma_1 + \sigma_2)^2 \tag{2}$$

$$\tau = (2\epsilon_1' + \epsilon_2')/4\pi(2\sigma_1 + \sigma_2). \tag{3}$$

If, instead of a dilute suspension of spheres, one takes as model a two-layered dielectric composed of sheets of two electrically different materials with the same thickness, the form of equation (1) is unchanged. The values of k and τ in equations (2) and (3) are changed only numerically by this drastic change in geometry. Experimental confirmation of the relative insensitivity of the Maxwell-Wagner dispersion to the geometric distribution of phases has been given by studies on a model system composed of quartz and alcohol.[18] The essential prerequisite for the Maxwell-Wagner dispersion is therefore the existence of boundaries between media of different electrical properties. Consequently we shall assume that the macroions and their associated counter ions may be represented by spherical droplets, although the actual geometry must tend toward cylindrical as dilution increases. Admittedly, the particles of such a pseudo-suspension would have no sharp boundary, but a fair approximation might be made by using a step function, defining as dispersed phase the regions containing polycations plus all nearby gegenions of potential energy nkT (where n is a small integer), and as continuous phase all the rest of the solution. Due to the higher concentration of ions in the dispersed phase, the latter would have a higher conductance than the bulk of the solution; the necessary condition for the existence of a Maxwell-Wagner dispersion is thus satisfied.

In the case of dilute polyelectrolyte solutions, we may set

$$\epsilon_1' = \epsilon_2' = \epsilon'$$

so that

$$k = 3p(\sigma_1 - \sigma_2)^2/(2\sigma_1 + \sigma_2)^2 \tag{4}$$

and

$$\tau = 3\epsilon'/4\pi(2\sigma_1 + \sigma_2). \tag{5}$$

The constant k is related to the zero frequency dielectric constant, ϵ_0, by the equation

$$\epsilon_0 = \epsilon_\infty(1 + k)$$

so that

$$k = (\epsilon_0 - \epsilon_\infty)/\epsilon_\infty. \tag{6}$$

We may use the experimental values of $\Delta\epsilon'$ at a frequency of 1 kilocycle as a crude measure of $\Delta\epsilon_0 = (\epsilon_0 - \epsilon_\infty)$ and thereby calculate effective values of p for various assumed values of (σ_2/σ_1) by means of equations (4) and (6). As may be seen from table 3, this procedure leads to values of p and (σ_2/σ_1) which are physically reasonable.

By the use of equation (5), we may calculate values of the critical frequency $f_c = 1/2\pi\tau$, using the known total bromide concentration, values of p from table 3, and various assumed values of (σ_2/σ_1). The resulting values of f_c are given in table 4. It will be seen that the agreement with the experimental curves of figure 1 is quite good. The effects of increased polymer concentration and of added simple salt may be understood in terms of equations (4) and (5). In either case, the conductivity is increased and the difference in conductivity between the dispersed phase and the continuous medium is decreased; this leads to a decreased dielectric increment per gram and an increased critical frequency.

TABLE 3

CALCULATION OF p AS FUNCTION OF (σ_2/σ_1)

		p		
g	$\Delta\epsilon_0$	$(\sigma_2/\sigma_1) = \infty$	10	3
0.050	40	0.17	0.30	1.05
0.016	17	0.07	0.12	0.44
0.005	7	0.03	0.05	0.18
0.001	2	0.008	0.015	0.052

TABLE 4

CALCULATION OF CRITICAL FREQUENCY AS FUNCTION OF (σ_2/σ_1)

			f_c (KC.)		
g	10^6 [Br']	$10^6\ \sigma_0$	$(\sigma_2/\sigma_1) = \infty$	10	3
0.050	160	12.0	590	310	100
0.016	52	4.1	490	200	89
0.005	16	1.2	330	83	37
0.001	3.2	0.25	260	22	9

The above discussion has been in terms of a uniform migration of charge within each medium due to the applied electric field. Such a migration might be considered analogous to the mechanism proposed by Kirkwood and Shumaker,[2] except that in this case the distribution of sites is a continuum. An alternative explanation of our results could be the relaxation effects in the ionic atmospheres of the polycations. Debye and Falkenhagen[19] showed that the conductance of simple electrolytes exhibits a dispersion at frequencies corresponding to the time of relaxation of the ion atmosphere. Such effects would presumably be enhanced in our case due to the high effective charge density of the polyelectrolyte. A dispersion in conductance means that κ, the specific conductance, is a complex quantity, and the imaginary term of $\kappa = (\kappa' - j\kappa'')$ represents an effective increment in dielectric constant (just as the loss factor ϵ'' in the complex dielectric constant $\epsilon = \epsilon' - j\epsilon''$ can be represented as an equivalent parallel conductance). The fact that our observed dispersion region moves to lower frequencies with increasing dilution (i.e., with increasing dimensions of ionic atmospheres, and hence longer time constants) is in agreement with this mechanism.

Whether the observed increments in dielectric constant are primarily due to a Maxwell-Wagner dispersion or primarily to a Debye-Falkenhagen effect, or quite possibly to a combination of the two, cannot be decided on the basis of the facts available at present. The shape of the dispersion curves of figures 1 and 3 does indeed suggest the presence of two dispersion ranges, indicating two relaxation times (or two most probable values if a distribution of relaxation times is involved). Studies must be made on more homogeneous polymers, however, before this can be taken as definite evidence for the existence of two mechanisms. Finally, of course, we must not exclude the possibility that some hitherto unsuspected effect is operating in these systems.

* Acknowledgment is made to the Office of Naval Research and to the National Institutes of Health for joint support of the work reported in this paper.

[1] Oncley, J. L., Chapter 22 in *Proteins, Amino Acids and Peptides* by Cohn, E. J., and Edsall J. T., Reinhold, New York (1943).

[2] Kirkwood, J. G., and Shumaker, J. B., these PROCEEDINGS, **38,** 855 (1952).

[3] Jungerer, G., *Acta Physiol. Scand.*, **10,** Supp. 32 (1945).

[4] Jungerer, I., *Ibid.*, **20,** Supp. 69 (1950).

[5] Allgén, L. G., *Ibid.*, **22,** Supp. 76 (1950).

[6] Oncley, J. L., Dintzis, H. M., and Fuoss, R. M., *Science*, **117,** 470 (1953).

[7] Allgén, L. G., and Roswall, S., *J. Polymer Sci.*, in press (1954).

[8] Oncley, J. L., *J. Am. Chem. Soc.*, **60,** 1115 (1938); *J. Phys. Chem.*, **44,** 1103 (1940).

[9] Hollies, N. R. S., and Oncley, J. L., Conference on Electrical Insulation, National Research Council (1950).

[10] Ferry, J. D., and Oncley, J. L., *J. Am. Chem. Soc.*, **63,** 272 (1941).

[11] Bailey, F. E., Patterson, A., Jr., and Fuoss, R. M., *Ibid.*, **74,** 1845 (1952).

[12] Fuoss, R. M., and Strauss, U. P., *Ann. N. Y. Acad. Sci.*, **51,** 836 (1949).

[13] Fuoss, R. M., and Signer, R., *J. Am. Chem. Soc.*, **73,** 5872 (1951).

[14] Fuoss, R. M., with Katchalsky, A., and Lifson, S., these PROCEEDINGS, **37,** 579 (1951).

[15] Maxwell, J. C., *A Treatise on Electricity and Magnetism*, Vol. II, 3rd ed., Oxford, 1892, p. 440

[16] Wagner, K., *Ann. Physik*, **40,** 817 (1913).

[17] Gemant, A., *Elektrophysik der Isolierstoffe*, Springer, Berlin, 1930, p. 105.

[18] Race, H. H., *Trans. Am. Inst. Elec. Engrs.*, **52,** 682 (1933).

[19] Debye, P., and Falkenhagen, H., *Physik. Z.*, **29,** 121, 401 (1928).

---

*THE RATES OF REACTION BETWEEN ALKYL HALIDES AND TERTIARY AMINES AS DETERMINED CONDUCTOMETRICALLY**

BY ROBERT P. LARSEN AND CHARLES A. KRAUS

METCALF RESEARCH LABORATORY, BROWN UNIVERSITY

Communicated December 7, 1953

The alkyl halides react quantitatively and more or less readily with tertiary amines at ordinary temperatures. The product of reaction is a quaternary ammonium salt which normally is soluble in the solvent in which the reaction is carried out.

Menschutkin[1] was the first to study the rate of this reaction some sixty years ago. He investigated the rate of reaction between ethyl iodide and triethylamine in 19 different solvents at 100°C. Although numerous investigators have studied reactions of this type since Menschutkin's time, accurate data are lacking for re-

action rates and we have little knowledge of the influence of various constitutional factors of reactants and of solvent on the rates of reaction.

It seemed worth while to devise a more precise method for determining the rate of these reactions and to determine the rate constants for a number of typical reactions. Since the quaternary ammonium salts are strong electrolytes, it seemed that a conductometric method might be successfully employed. Preston and Jones[2] employed this method earlier to measure the reaction rates between dimethylaniline and allyl and methyl iodides in ethanol at 25°. They found the rate constants determined conductometrically to agree well with those determined gravimetrically. However, they failed to take into account several sources of error as will appear below.

The conductance of the pure salt, the product of the reaction to be studied, was first measured over a concentration range that included salt concentrations as produced in the course of reaction. In general, the conversion of reactants to salt was less than 0.1%. In most cases, the concentration of reactants varied between 0.01 and 0.1 *N*. For a few reactions, where the reaction rate was low, the concentration of reactants varied between 0.1 and 1.0 *N*.

TABLE 1

TYPICAL RATE DETERMINATION FOR REACTIONS OF ETHYL IODIDE AND TRIETHYLAMINE IN ACETONE

C, EtI = 3.8×10^{-2}		*C*, $Et_3N = 3.752 \times 10^{-2}$		D = 0.788
TIME, SEC.	RESISTANCE, OHMS	SPECIFIC CONDUCTANCE $\times 10^5$	PRODUCT CONCENTRATION $\times 10^3$	$\Delta C/\Delta t$ $\times 10^6$
176	40,000	4.00	1.99	...
309	24,000	6.66	3.31	9.93
422	18,000	8.88	4.42	9.89
552	14,000	11.41	5.72	9.91
683	11,500	13.89	6.97	9.81
810	9,800	16.30	8.22	9.81
943	8,500	18.79	9.54	9.86
1078	7,500	21.30	10.85	9.81
1205	6,750	23.66	12.09	9.81

With reactions of the trialkylamines in acetone, where reactant concentrations were low, it was unnecessary to apply a correction for the effect of the reactants on the conductance of the reaction product. For the reactions with pyridine, the conductance of the salt was measured in the presence of varying amounts of each of the reactants. In applying the correction for this effect, it was assumed that the conductance change due to the presence of both reactants was equal to the sum of the effects of the two reactants when present alone. Solutions of ethyl iodide in the alcohols are not entirely stable, as Norris and Prentiss[3] have shown for butyl bromide in butanol. Solvolysis takes place with the formation of the halogen acid and the conductance of the solution increases slowly. The rate of this change was measured for solutions of ethyl iodide in methanol and ethanol and corrections were made to the observed conductance values of the reaction mixtures.

Conductance measurements of the reaction mixtures, as also of the pure salts, were carried out in a cell of the Erlenmeyer type[4] having a capacity of 125 cc. with unplatinized electrodes. The measurements were carried out at 25 ± 0.01°C. Resistances were measured with a precision bridge having a capacity of 100,000 ohms.

Weighed quantities of solvent and amine were introduced into the conductance cell, the alkyl halide was added, and after mixing the cell was placed in a thermostat, the resistance of the mixture measured, and the time noted. A new resistance was introduced into the bridge circuit and the time was noted when balance was

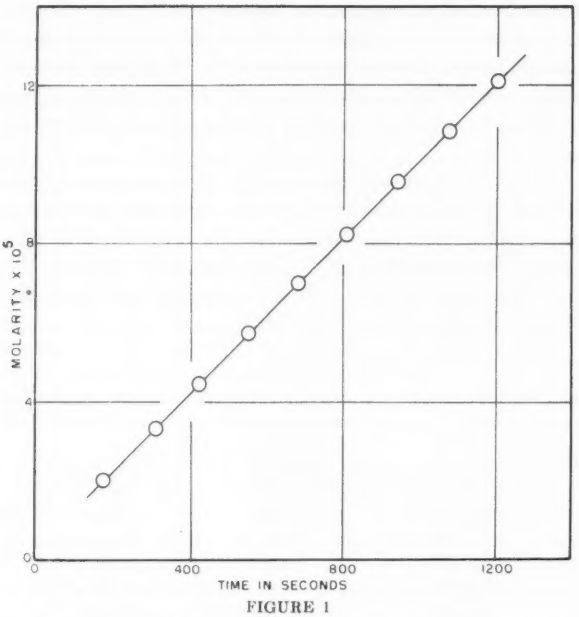


FIGURE 1

Rate plot for reaction of ethyl iodide and triethyl amine in acetone at 25°.

again reached. Thereafter, a new adjustment of resistance was made and this procedure was repeated until the desired amount of reaction had taken place. The cell was then weighed to determine the amount of alkyl halide and the density of the solution was determined with a Westphal balance.

The equation for the reaction rate reads:

$$dx/dt = KC_aC_b, \tag{1}$$

where x is the moles of product per liter of solution, t is the time in seconds, K is the rate constant, and C_a and C_b are the concentrations of halide and amine. Since the maximum concentration change of the reactants is of the order of 0.1%, x is a linear function of t within the experimental error. A plot of the concentration of reaction product vs. t should be linear and its slope should yield the value of dx/dt, wherefrom the value of the reaction constant K may be computed.

For the purpose of illustration, details for one reaction between ethyl iodide and triethylamine are given in table 1. At the head of the table are given values of the concentration of the two reactants and the density of the solution. Time appears in column 1, the resistance of the solution in column 2, its specific conductance in

column 3, and the concentration of the reaction product in column 4. In the last column, are given values of the rate dx/dt in moles per second for the time intervals as listed in column 1. There is some variation of rate for the earliest points, possibly due to lack of thermal equilibrium.

TABLE 2

CONSTANTS FOR REACTIONS OF ALKYL HALIDES WITH TRIETHYLAMINE

C HALIDE $\times 10^2$	C AMINE $\times 10^2$	$K \times 10^5$	C HALIDE $\times 10^2$	C AMINE $\times 10^2$	$K \times 10^5$
	A. EtI + Et_3N			B. EtI + Pr_3N	
2.592	2.051	6.98	3.84	3.28	1.53
1.541	4.751	6.82	4.68	4.82	1.51
2.439	4.277	6.84	5.32	5.68	1.51
3.840	3.752	6.81	4.50	6.76	1.50
5.77	8.17	6.57	7.08	7.65	1.48
9.21	11.41	6.60			
	C. EtI + Bu_3N			D. EtI + Am_3N	
1.930	2.701	2.03	1.85	2.32	2.04
3 772	2.996	2.03	2.06	2.66	2.10
4.85	3.69	2.09	3.09	2.57	2.02
6.01	4.64	2.09	4.00	4.01	2.02
5.20	6.71	2.07			
	E. BuI + Pr_3N			F. PrI + Pr_3N	
4.88	5.01	0.238	4.40	5.64	0.276
5.69	4.47	0.242	4.94	5.39	0.277
5.36	6.46	0.243	6.49	8.40	0.273
4.74	7.12	0.240	7.02	8.86	0.269
	G. PrBr + Pr_3N			H. PrCl + Pr_3N	
6.97	4.80	0.137	5.02	6.97	1.34[a]
9.81	5.61	0.139	9.83	6.48	1.22[a]
8.59	7.44	0.134	13.30	6.06	0.90[a]

[a] $K \times 10^9$.

TABLE 3

REACTIONS OF ETHYL IODIDE AND PYRIDINE IN DIFFERENT SOLVENTS

C EtI $\times 10$	PYRIDINE $\times 10$	$K \times 10^6$	C EtI $\times 10$	PYRIDINE $\times 10$	$K \times 10^6$	C EtI $\times 10$	PYRIDINE $\times 10$	$K \times 10^6$
	A. Acetone			B. Methanol			C. Ethanol	
0.938	0.735	9.24	2.76	2.58	1.20	1.852	1.584	0.776
0.923	0.798	9.33	3.06	2.49	1.19	2.483	2.627	0.850
0.965	0.926	9.21	2.94	3.21	1.20	3.91	3.12	0.882
1.227	1.290	9.34	3.94	3.58	1.22	3.27	3.56	0.885
1.318	1.155	9.19	4.45	3.82	1.21	3.28	4.00	0.916
			4.89	4.90	1.24	3.60	5.16	0.966
						4.92	5.75	0.984
						8.09	9.29	1.178

A plot of the results is shown in figure 1. The value of dx/dt was found to be 9.82×10^{-6} moles per second. This leads to value of 6.81×10^{-5} for the rate constant K. As stated above, it was unnecessary to make a correction for the effect of the reactants on the conductance of the salt for the aliphatic amines in acetone.

All the materials employed were carefully purified. The tertiary amines[5] were converted to picrates, recrystallized, and reconverted to amine. Pyridine was a pure product supplied to us by Dr. C. J. Carignan, of this laboratory.[6] Alkyl halides were treated with dilute caustic to remove halogens and redistilled after drying. Acetone[7] was purified by the method of Reynolds.[8] The final drying of methanol and ethanol was carried out with activated alumina. The specific conductance of the acetone ranged between 1 and 2×10^{-9}, that of methanol between 1 and 0.5×10^{-7}, and that of ethanol between 1 and 0.5×10^{-8}.

The results of the present investigation are given in tables 2 and 3, where the concentration of alkyl halide appears in column 1, that of amine in column 2, and the rate constant K in column 3. Mean values of the rate constants are summarized in table 4.

With the exception of two reactions, the rate constant is not dependent on reactant concentration in the concentration ranges employed. For different series of reactions with given reactants, the average deviation of K from the mean is of the order of 1.1% with a maximum deviation of 1.7% for one reaction.

TABLE 4

RATE CONSTANTS FOR REACTIONS OF ALKYL HALIDES WITH TERTIARY ANIONS IN DIFFERENT SOLVENTS

REACTANTS	RATE CONSTANT	SOLVENT
EtI + Et_3N	6.77×10^{-5}	acetone
EtI + Pr_3N	1.51×10^{-5}	acetone
EtI + Bu_3N	2.05×10^{-5}	acetone
EtI + Am_3N	2.04×10^{-4}	acetone
BuI + Pr_3N	2.41×10^{-6}	acetone
PrI + Pr_3N	2.74×10^{-6}	acetone
PrBr + Pr_3N	1.37×10^{-6}	acetone
PrCl + Pr_3N	1×10^{-9}	acetone
EtI + Pyr	9.26×10^{-6}	acetone
EtI + Pyr	1.21×10^{-6}	methanol
EtI + Pyr	7×10^{-7}	ethanol

The value of K for ethyl iodide and pyridine in ethanol increases markedly with increasing reactant concentration. In contrast, the constant for the same reaction in methanol and in acetone is independent of concentration. Whether the change of K in ethanol is due to some unknown source of error or not must remain uncertain. The alkyl halides are not stable in ethanol and correction had to be made for this effect. However, the same is true of methanol and there the constant exhibits no dependence on reactant concentration.

The constant for the reaction between propyl chloride and tripropylamine increases markedly with concentration. Here, the reaction was very slow, the concentration of reactants was high and time intervals amounted to several hours. The concentration of the product of reaction ranged between 0.6 and 1.0×10^{-5} N. Under these circumstances various sources of error might creep into the observations. It may safely be concluded, however, that the rate constant of this reaction is of the order of 1×10^{-9}.

It is of interest to compare the rate of the reactions between tripropylamine and propyliodide, bromide, and chloride. The constants for these reactions are, respectively, 2.74×10^{-6}, 1.37×10^{-6}, and 1×10^{-9}. While the rate constant of

the iodide is only twiee as great as that of the bromide, it is two thousand times as great as that of the chloride.

For the reaction of ethyl iodide with tertiary amines, the rate constant decreases greatly in going from triethyl to tripropylamine. The constants for tributyl- and triamylamine have practically the same value and both are markedly greater than that of tripropylamine, 2.05×10^{-5} and 1.51×10^{-5}, respectively. It is difficult to account for this inversion. Conceivably, one or another of the amines was impure; if so, it must have been the propyl derivative. However, this amine was carefully purified and there is no reason for doubting its purity other than the inversion under discussion.

As the alkyl group attached to the iodine atom becomes larger, the rate decreases. Thus, for reactions of ethyl, propyl, and butyl iodides with tripropylamine, the rates are, respectively, 1.51×10^{-5}, 2.74×10^{-6}, and 2.41×10^{-6}. The change from propyl to butyl is relatively small.

The effect of solvent on reaction rates is of interest. For the ethyl iodide-pyridine reaction, the rate constant for acetone (9.26×10^{-6}) is markedly greater than it is for methanol (1.21×10^{-6}) which, in turn, is higher than for ethanol. As already mentioned, the constant for ethanol is dependent on the concentration of reactants; extrapolating to low reactant concentrations, one obtains the approximate value 7×10^{-7}. Norris and Prentiss[3] have determined the rates for the same reaction in these same solvents. For acetone, methanol, and ethanol, they report the values 10×10^{-6}, 1.94×10^{-6}, and 1.08×10^{-6}, respectively. The values are in good agreement with those of the present investigation considering the much higher concentration of reactants which Norris and Prentiss employed. If we interpolate a value of K for ethanol at a concentration corresponding to that of Norris and Prentiss, we find $K = 1.07 \times 10^{-6}$ in good agreement with the earlier value.

There appears to be little or no correlation between the physical properties of the solvent and the rate of reaction. The dielectric constant of methanol is markedly greater than that of acetone, yet the rate in acetone is much greater than in methanol. The viscosity of acetone is less than that of methanol but not greatly so. Menschutkin[1] measured the reaction rates of ethyl iodide and triethylamine in 19 different solvents and found the reaction rate to vary widely for different solvents. While Menschutkin's measurements are subject to rather large uncertainties, they nevertheless, show that the rate of reaction is greatly dependent on the solvent medium. Further studies of reaction rates in different solvents would seem to be called for. Conceivably such studies would throw some light on the detailed mechanism which determines the reaction rate and in what manner such mechanism is dependent upon constitutional factors of solvent as well as of reactants. In carrying out precise measurements, the conductometric method may be recommended.

[4] This paper is based on a portion of a thesis presented by Robert P. Larsen in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Brown University, May, 1948.

[1] Menschutkin, N., *Z. phys. Chem.*, **6**, 41 (1890).

[2] Preston, R. W. D., and Jones, H. O., *J. Chem. Soc.*, **101**, 1950 (1912).

[3] Norris, J. F., and Prentiss, S. W., *J. Am. Chem. Soc.*, **50**, 3042 (1928).

[4] Fuoss, R. M., and Kraus, C. A., *Ibid.*, **55**, 21 (1933).

[5] We are indebted to the Sharples Chemicals, Inc., for supplying the amines used in this investigation.

[6] Carignan, C. J., and Kraus, C. A., *J. Am. Chem. Soc.*, **71**, 2983 (1949).
[7] We are indebted to the Tennessee Eastman Company for supplying us with this solvent.
[8] Reynolds, M. B., and Kraus, C. A., *J. Am. Chem. Soc.*, **70**, 1709 (1948).

SONIC POINT ON SUPERSONIC AIRFOILS

BY T. Y. THOMAS

INSTITUTE FOR NUMERICAL ANALYSIS, LOS ANGELES, CALIFORNIA

Communicated December 4, 1953

1. Introduction.—In figure 1 we have indicated an airfoil in a uniform supersonic stream of velocity W. Under the conditions assumed the shock AQ will be detached from the airfoil and the stream line AB, whose continuation is the contour of the airfoil, will be parallel to the direction of the uniform flow and perpendicular to the shock at the point A. The sonic line, i.e., the curve along which the mach number $m = 1$, is represented by PQ. This line divides the region of the flow behind the shock into a subsonic ($m < 1$) region I and a supersonic ($m > 1$) region II. It is the object of this note to determine the position of the sonic point P on the airfoil.[1]

Introducing the entropy S as a parameter along the sonic line we can write

$$\omega_P = \omega_Q + \left(\frac{d\omega}{dS}\right)_R (S_0 - S_Q), \tag{1}$$

where ω denotes the inclination of the stream lines with ω_P and ω_Q the inclinations of these lines at points P and Q respectively. Also S_0 is the entropy along the stream line $ABPC$, and hence is the entropy at the point P, while S_Q is the entropy at Q. The derivative $d\omega/dS$ in the above equation is evaluated at some point R between P and Q on the sonic line. Approximately we have

$$\omega_P = \omega_Q + \left(\frac{d\omega}{dS}\right)_Q (S_0 - S_Q), \tag{2}$$

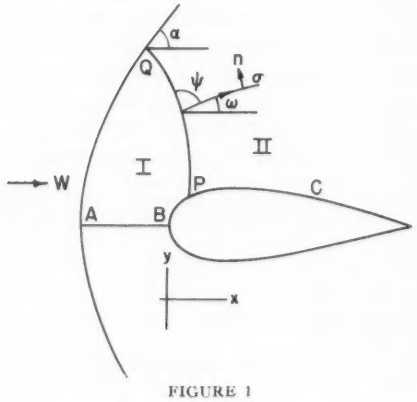


FIGURE 1

where the derivative is now evaluated at the point Q. The expression (2) for ω_P will be correct to within terms of the second order in the entropy difference $S_0 - S_Q$ and for flows in which the entropy changes are small it may be expected that the relation (2) will give a good approximation to the inclination ω_P of the airfoil at the sonic point P.

The remainder of this paper is devoted to the derivation of the formulas for the quantities in the right member of (2). It is assumed in this discussion that we are dealing with an ideal gas

and that viscosity and thermal conductivity can be neglected. Computations[2] are carried out in the last section and in this section we have also shown the graph of the inclination ω_P as a function of the mach number M of the uniform stream.

2. Determination of ω_Q.—The inclination of the stream line at the point Q behind the shock is determined directly from the condition that $m = 1$ at this point. In fact we have[3]

$$\sin^2 \alpha = \frac{1}{4\gamma M^2}\left\{[(\gamma+1)M^2 + (\gamma-3)] + \sqrt{[(\gamma+1)M^2 + (\gamma-3)]^2 + 16\gamma}\right\}, \tag{3}$$

where α is the inclination of the shock line at Q and γ is a constant having the value 1.405 for the case of air flow with which we shall be concerned. From (3) we find the value of $\alpha(<90°)$ at Q. The inclination $\omega = \omega_Q$ is then given by the shock relation

$$\tan \omega = \frac{2[(M^2-1)\tan^2 \alpha - 1]}{[(\gamma-1)M^2+2]\tan^3 \alpha + [(\gamma+1)M^2+2]\tan \alpha}. \tag{4}$$

3. Inclination of the Sonic Line Relative to the Stream Lines.—The angle ψ in figure 1 gives the inclination of the sonic line relative to the stream lines. Now from the fact that $m = \text{const.}$ on the sonic line it follows readily that the flow velocity v is likewise constant on this line. Letting l denote arc length along the sonic line measured from P to Q we therefore have

$$\frac{dv}{dl} = \frac{dv}{d\sigma}\cos\psi + \frac{dv}{dn}\sin\psi = 0,$$

where σ is the arc length along the stream lines in the direction of the flow and n denotes the direction normal to the stream lines as shown in figure 1. Hence

$$\frac{dv}{dn}\tan\psi = -\frac{dv}{d\sigma} = \frac{1}{\rho v}\frac{dp}{d\sigma},$$

where ρ, as usual, denotes the density. Let us now put $p = p(\omega, S)$, i.e., we take ω and S as the independent variables in the representation of the pressure function. Substituting for the derivative dv/dn the above equation can then be written

$$(vK + Ap)\tan\psi = \frac{K}{\rho v}\frac{\partial p}{\partial \omega},$$

where K is the curvature of the stream lines and A is an invariant of the flow which is constant along stream lines. Or, we have

$$\gamma\left[\frac{v^2}{\gamma p/\rho} + \frac{pv^2}{\gamma p/\rho}\frac{A}{vK}\right]\tan\psi = \frac{1}{p}\frac{\partial p}{\partial \omega}.$$

Hence putting $c^2 = \gamma p/\rho$ where c is the velocity of sound and $m^2 = v^2/c^2 = 1$ we obtain

$$\tan\psi = \frac{\partial p/\partial\omega}{\gamma p(1+D)}, \qquad D = \frac{pA}{vK}. \tag{5}$$

Explicit formulas are given in §6 for the quantity D and in §7 for the quantity $(\partial p/\partial\omega)/p$ at the point Q. Calculations based on these formulas, which have not been entered in this note, indicate that $\tan\psi$ can have both positive and negative values at Q depending on the mach number M and that this function will become infinite only in the exceptional case for which $D = -1$. Since the relation (5) is exact, computations are exact, i.e., it is possible to determine to any degree of approximation the inclination ψ of the sonic line relative to the stream line at the point Q where the sonic line meets the shock.

4. Formula for $d\omega/dS$.—It can be shown that[4]

$$(m^2 - 1)\frac{dp}{d\sigma} = -\rho v^2 \frac{d\omega}{dn},$$

$$\frac{dS}{dn} = -(\gamma - 1)\rho v A.$$

From the first of these formulas we have $d\omega/dn = 0$ along the sonic line under the assumption that $dp/d\sigma$ does not become infinite; this quantity is definitely finite at the point Q, with which we shall primarily be concerned in this note, in consequence of the following equation (24). Hence along the sonic line we have

$$\frac{d\omega}{dl} = \frac{d\omega}{d\sigma}\frac{d\sigma}{dl} + \frac{d\omega}{dn}\frac{dn}{dl} = K\cos\psi,$$

$$\frac{d\omega}{dl} = \frac{d\omega}{dS}\frac{dS}{dn}\frac{dn}{dl} = -(\gamma - 1)\rho v A \sin\psi \frac{d\omega}{dS}.$$

The combination of these two equations now gives

$$\frac{d\omega}{dS} = \frac{-1}{(\gamma - 1)\rho v\left(\frac{A}{K}\right)\tan\psi}. \tag{6}$$

But,

$$\frac{\rho v A}{K} = \left(\frac{pA}{vK}\right)\frac{\gamma v^2}{\gamma p/\rho} = D\frac{\gamma v^2}{c^2} = \gamma D$$

along the sonic line. Hence (6) becomes

$$\frac{d\omega}{dS} = \frac{-1}{\gamma(\gamma - 1)D\tan\psi}. \tag{7}$$

5. The Curvature Ratio G_0.—Between the curvature K of the stream line and the curvature κ of the shock line we have the relation $K = G_0\kappa$ at any point behind the shock, where[5]

$$-G_0 = \frac{\left(\frac{\rho}{\rho_1} + \frac{\rho_1}{\rho} + \frac{6 - 2\gamma}{\gamma + 1}\right)\left(\frac{u_t}{v}\right)^3 - \left(\frac{\rho}{\rho_1} - 1\right)\left(\frac{u_t}{v}\right)}{(u_n^2/c^2) - 1} + \frac{4}{\gamma + 1}\left(\frac{u_t}{v}\right). \tag{8}$$

Here the subscript 1 is used to denote a quantity in the uniform incident flow.

The tangential velocity component is specified by u_t and u_n denotes the normal component of velocity behind the shock. Other quantities appearing in (8) have already been defined.

At the point Q where $m = 1$ it follows readily that

$$(u_n^2/c^2) - 1 = -(u_t/c)^2. \tag{9}$$

Making this substitution in (8) this equation can be reduced to

$$-G_0 = \left[\frac{2(\gamma - 1)}{\gamma + 1} - \left(\frac{\rho}{\rho_1} + \frac{\rho_1}{\rho}\right)\right]\left(\frac{u_t}{v}\right) + \frac{[(\rho/\rho_1) - 1]}{u_t/v}. \tag{10}$$

Now

$$[\rho]/\rho_1 = (\rho - \rho_1)/\rho_1 = (\rho/\rho_1) - 1, \tag{11}$$

$$[\rho]/\rho = (\rho - \rho_1)/\rho = 1 - (\rho_1/\rho), \tag{12}$$

and hence

$$[\rho]^2/\rho\rho_1 = \rho/\rho_1 + (\rho_1/\rho) - 2. \tag{13}$$

Using (11) and (13) equation (10) becomes

$$G_0 = \left[\frac{4}{\gamma + 1} + \frac{[\rho]^2}{\rho\rho_1}\right]\left(\frac{u_t}{v}\right) - \frac{[\rho]/\rho_1}{u_t/v}. \tag{14}$$

We now make use of the following shock condition[6]

$$[\rho] = \frac{-2\rho(\rho u_n^2 - \rho p)}{2\gamma p + (\gamma - 1)\rho u_n^2}. \tag{15}$$

Applying this condition at the point Q we have

$$[\rho] = \frac{-2\rho(u_n^2 - c^2)}{2c^2 + (\gamma - 1)u_n^2} = \frac{-2\rho[(u_n^2/c^2) - 1]}{2 + (\gamma - 1)u_n^2/c^2},$$

or,

$$\frac{[\rho]}{\rho} = \frac{2(u_t/v)^2}{2 + (\gamma - 1)[1 - (u_t/v)^2]} = \frac{2(u_t/v)^2}{(\gamma + 1) - (\gamma - 1)(u_t/v)^2}. \tag{16}$$

Now we can write

$$\frac{\rho}{\rho_1} = \frac{1}{1 - [\rho]/\rho},$$

and hence

$$\frac{[\rho]}{\rho_1} = \frac{[\rho]}{\rho}\frac{\rho}{\rho_1} = \frac{[\rho]/\rho}{1 - [\rho]/\rho}. \tag{17}$$

When we replace the quantity $[\rho]/\rho$ in (17) by its value as given by (16) we obtain the following relation

$$\frac{[\rho]}{\rho_1} = \frac{2a^2}{(\gamma + 1)(1 - a^2)}, \tag{18}$$

where we have put $a = u_t/v$. Also we can write

$$\frac{[\rho]}{\rho} = \frac{2a^2}{(\gamma + 1) - (\gamma - 1)a^2} \tag{19}$$

from (16). Now when we make the substitutions (18) and (19) in (14) we find, after some reduction, the following relation

$$G_0 = \frac{2a}{\gamma + 1}\left[\frac{2(\gamma - 1)a^2 - (\gamma + 1)}{(\gamma - 1)a^2 - (\gamma + 1)}\right]. \tag{20}$$

Finally replacing a by its value $-\cos(\alpha - \omega)$ equation (20) can be given the form

$$G_0 = \frac{-2\cos(\alpha - \omega)}{\gamma + 1}\left[\frac{(\gamma + 1)\tan^2(\alpha - \omega) + 3 - \gamma}{(\gamma + 1)\tan^2(\alpha - \omega) + 2}\right]. \tag{21}$$

6. The Ratio A/K.—It has been shown that[7] $A/K = -[\rho]^2 u_t/\rho_1 \rho p G_0$ at points behind the shock. Hence from the second equation (5) we have

$$D = \frac{-[\rho]^2}{\rho_1 \rho G_0}\left(\frac{u_t}{v}\right) = \frac{[\rho]^2}{\rho_1 \rho G_0}\cos(\alpha - \omega). \tag{22}$$

At the point Q on the sonic line we can make the substitutions (18), (19), and (21) in the right member of (22). After some reduction we thus obtain

$$D = -2/\tan^2(\alpha - \omega)[(\gamma + 1)\tan^2(\alpha - \omega) + 3 - \gamma]. \tag{23}$$

7. Formula for $\partial p/\partial\omega$.—At any point behind the shock line we have[8]

$$\frac{\partial p}{\partial \omega} = -\rho v^2\left(\frac{u_n}{u_t}\right) - \frac{4\rho v u_n}{(\gamma + 1)G_0}.$$

Since $u_n = v\sin(\alpha - \omega)$ and $u_t = -v\cos(\alpha - \omega)$ we furthermore obtain

$$\frac{1}{p}\frac{\partial p}{\partial \omega} = \frac{\rho v^2}{p}\left[\tan(\alpha - \omega) - \frac{4}{\gamma + 1}\frac{\sin(\alpha - \omega)}{G_0}\right].$$

Hence at the point Q we have

$$\frac{1}{p}\frac{\partial p}{\partial \omega} = \gamma\left[\tan(\alpha - \omega) - \frac{4}{\gamma + 1}\frac{\sin(\alpha - \omega)}{G_0}\right]$$

and then making the substitution (21) for G_0 we obtain an equation which can be written in the form

$$\frac{1}{p}\frac{\partial p}{\partial \omega} = \gamma\tan(\alpha - \omega)\left[\frac{3(\gamma + 1)\tan^2(\alpha - \omega) + 7 - \gamma}{(\gamma + 1)\tan^2(\alpha - \omega) + 3 - \gamma}\right]. \tag{24}$$

8. The Entropy Difference $S_0 - S_Q$.—The entropy S in the non-isentropic flow behind the shock is defined by

$$S = \log\frac{N}{N_1} = \log\frac{p/p_1}{(\rho/\rho_1)^\gamma}.$$

At points immediately behind the shock line the ratios p/p_1 and ρ/ρ_1 are determined

from the shock conditions as follows

$$p/p_1 = (p/p_1)_* = \frac{2\gamma(M^2 \sin^2 \alpha - 1)}{\gamma + 1} + 1,$$

$$\rho/\rho_1 = (\rho/\rho_1)_* = \frac{2(M^2 \sin^2 \alpha - 1)}{(\gamma - 1)M^2 \sin^2 \alpha + 2} + 1.$$

By selecting the value of α given by (3) we can therefore find the entropy S_Q at the point Q on the sonic line.

The entropy S_0 along the contour of the airfoil is found by taking $\alpha = 90°$ in the above expressions for p/p_1 and ρ/ρ_1 since the shock is assumed to be perpendicular to the stream line $ABPC$ at the point A (see figure 1). Hence if we denote the values of the pressure and density ratios at A by p_v/p_1 and ρ_v/ρ_1 we have

$$S_0 = \log \frac{p_v/p_1}{(\rho_v/\rho_1)^\gamma},$$

where

$$\frac{p_v}{p_1} = \frac{2\gamma(M^2 - 1)}{\gamma + 1} + 1,$$

$$\frac{\rho_v}{\rho_1} = \frac{2(M^2 - 1)}{(\gamma - 1)M^2 + 2} + 1.$$

The above relations permit the calculation of the entropy difference $S_0 - S_Q$ which occurs in the equation (2).

9. Computations and Determination of the Sonic Point on the Airfoil.—The method of determining the quantity ω_Q is given in §2 and the determination of the entropy difference $S_0 - S_Q$ is explained in §8. The determination of the derivative $d\omega/dS$ at Q which also occurs in the relation (2) is covered by the formulas of the preceding sections. Thus the quantity D is obtained from (23) when the value of α has been found from (3) and $\omega = \omega_Q$ from (4). These values of α and ω_Q likewise permit the determination of the quantity in the left member of (24) after which we obtain $\tan \psi$ from equation (5). With $\tan \psi$ and D determined we find the value of the derivative $d\omega/dS$ at Q from equation (7). Hence we can determine the variation

$$\delta\omega = \left(\frac{d\omega}{dS}\right)_Q (S_0 - S_Q)$$

in the inclination ω and hence the value of ω_P by equation (2).

In table 1 we have given the values of the quantities ω_Q, $\delta\omega$ and ω_P for various values of the mach number M of the incident flow. Figure 2 shows the graphical relationship between ω_P and M.

Remark 1. The representation of the mach number m as a function $m(\omega, S)$ of the independent variables ω and S suggests the possibility of taking $m = m(\omega)$ as a first approximation of this quantity where $m(\omega)$ is the function of ω obtained by evaluating the mach number behind the shock. This corresponds to the usual neglect of the entropy in gas dynamical problems. From this standpoint we are

led to the value of ω_Q, which is strictly the inclination of the stream line at the point where $m = 1$ behind the shock, as an approximation to the inclination of the contour at the sonic point P. The determination of the end point P of the sonic line by the value of ω_Q instead of ω_P as given by (2) would appear to be acceptable as an approximation from the fact that $\delta\omega$ is generally small in comparison with ω_Q.

TABLE 1

M	ω_Q	$\delta\omega$	ω_P
1.05	0.514	0.147	0.661
1.10	1.403	0.439	1.842
1.20	3.691	0.918	4.609
1.30	6.300	1.302	7.603
1.40	9.000	1.486	10.485
1.50	11.658	1.550	13.208
1.60	14.198	1.298	15.496
1.70	16.577	1.072	17.649
1.80	18.775	.920	19.695
1.90	20.790	.412	21.202
2.00	22.626	.059	22.685
2.50	29.554	−2.802	26.753
3.00	33.872	−5.261	28.612

Remark 2. The formula in §6 for the quantity A which is constant along stream lines can be written in the form

$$A = \frac{-[\rho]^2 u_t \kappa}{\rho_1 \rho p}.$$

Since u_t vanishes at the point A in figure 1 it follows that the quantity A and hence D in equation (5) will vanish along the contour of the airfoil. The inclination ψ of the sonic line at P relative to the contour is therefore given by

$$\tan \psi = \frac{\partial p/\partial \omega}{\gamma p} = \frac{dp/d\sigma}{\gamma p K},$$

where K is the curvature of the contour at P.

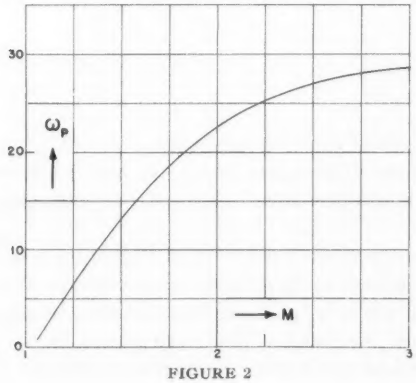


FIGURE 2

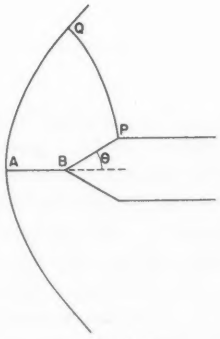


FIGURE 3

Remark 3. In the case of the wedge the end point P of the sonic line will lie at the shoulder[9] as shown in figure 3 on the basis of the discussion in this note provided that the angle ω_P as calculated from (2) is less than the wedge angle θ. For angles $\omega_P > \theta$, which will occur for detached shocks when the mach number M is sufficiently large, the point P will not be determined on the wedge by the method of this note. This raises the question as to whether the shock will not be attached to the vertex B of the wedge when $\omega_P > \theta$ so that the case of the detached shock under consideration does not apply. This is an interesting question and one which might well be the subject of a separate investigation. In considering this problem, however, it must be borne in mind that the equation (2) for the quantity ω_P is approximate and that the inequality $\omega_P > \theta$ must therefore be interpreted only as an approximate condition for the occurrence of the attached shock.

[1] Prepared for the Naval Research Laboratory, Washington, D. C.

[2] These computations were made on SWAC at the Institute for Numerical Analysis, Los Angeles, California.

[3] See T. Y. Thomas, "The Distribution of Singular Shock Directions," *Jour. of Math. and Physics*, **28**, p. 155 (1949).

[4] We have omitted the factor Jc_p from the second of these formulas since this quantity will cancel from the final results. See T. Y. Thomas, "The Determination of Pressure on Curved Bodies Behind Shocks," *Communications on Pure and Applied Mathematics, III*, p. 109 and p. 111 (1950).

[5] See T. Y. Thomas, "Calculation of the Curvatures of Attached Shock Waves," *Jour. of Math. and Physics*, **27**, p. 280 (1949).

[6] "The Distribution of Singular Shock Directions," *loc. cit.*, p. 164.

[7] "The Determination of Pressure on Curved Bodies Behind Shocks," *loc. cit.*, p. 112.

[8] See T. Y. Thomas, "Distribution of Pressure on Curved Profiles in Supersonic Gas Flow with Variable Entropy," these PROCEEDINGS, **36**, p. 112 (1950).

[9] The point P cannot lie at the vertex B of the wedge since the velocity and hence the mach number m will be zero at B for detached shocks. It may also be mentioned that boundary layer effects, which are neglected in this discussion, may actually play a significant role in the determination of the point P when the value of the angle ω_P is near to the value of the wedge angle θ.

REAR SHOCK AND PRESSURE ON SUPERSONIC AIRFOILS

BY W. F. BROWN* AND T. Y. THOMAS†

INSTITUTE FOR NUMERICAL ANALYSIS, LOS ANGELES, CALIFORNIA

Communicated December 4, 1953

1. Introduction.—Consider a curved airfoil along which the inclination ω is a monotonically decreasing function of the arc length with $\omega = 0$ at the point T (see figure 1). Let the curve $\omega =$ const. join the point P, situated between the vertex V and the point T, with the point Q on the shock which is formed when the airfoil is placed in a uniform supersonic stream. A (1, 1) correspondence $P \leftrightarrow Q$ between points on this shock line and the segment VT of the airfoil is thereby established such that $V \leftrightarrow V$ and $Q \rightarrow \infty$ as $P \rightarrow T$. This correspondence lies at the basis of the derivation of a simple formula giving an approximate determination of the pressure on the segment VT of the airfoil.[1] It is the primary purpose of this note to show that this formula can be extended to give the determination of pressure

on the remainder of the airfoil by an analogous consideration involving the shock which is formed at the rear of the airfoil. In this connection no attempt will be made to treat the problem of the separation of flow. The rear shock therefore has been drawn in figure 1 from the rear vertex R of the airfoil, as is frequently done, although this does not in general correspond to the physical fact.

In § 3 we have made the assumption, which is at best an approximation, that the flow behind the rear shock is identical with the uniform stream in which the airfoil is submerged. This assumption leads directly, on the basis of the shock conditions, to the approximation formula (7) for the pressure on the segment TR. It is seen that (7) is identical with the previous formula for pressure derived by a consideration of the shock conditions relative to the front shock and which therefore strictly applies to the determination of pressure on the segment VT. The discussion in § 3 confirms the extension of this previous pressure formula from the segment VT to the

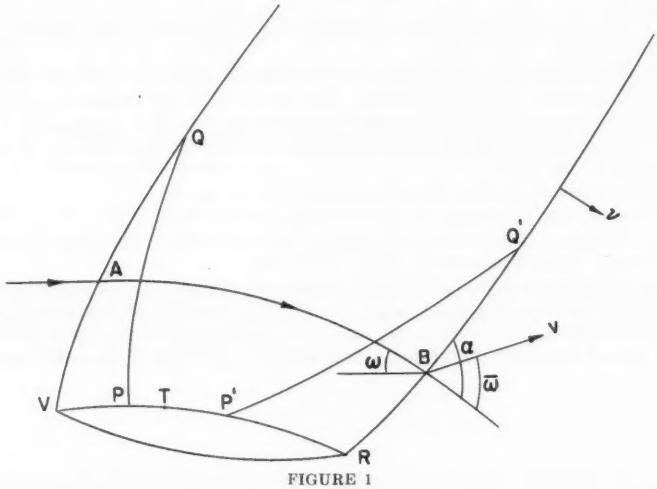


FIGURE 1

segment TR, or indeed of the more general formulae for pressure determination which are derived from considerations based on the front shock and which depend on entropy variations in the flow, since it appears unlikely that the flow behind the rear shock will differ greatly from the free stream.

It must be emphasized, however, that the use of these extended pressure formulae in no way implies that the flow quantities behind the rear shock are the same as those in the free stream in spite of the derivation in § 3. This is fortunate since it is shown in § 4 that the existence of the free stream flow behind the rear shock violates the entropy condition, i.e., the requirement that entropy cannot decrease across the shock front. Actually the entropy condition is also violated under the more general assumption of irrotational flow behind the rear shock as shown in § 5. Finally it is observed in § 6 that even the general rotational and non-isentropic flow behind the rear shock cannot be permitted in view of the *natural* boundary conditions, i.e., that the flow behind the rear shock approaches the free stream flow at infinity.

Influences such as viscosity and heat conduction, which are neglected in this discussion, may possibly allow the realization of these natural boundary conditions. It seems reasonable to assume, however, that for finite regions where viscosity and thermal conductivity are negligible the general non-isentropic theory will give a satisfactory solution of the flow problem.

2. The Shock Conditions.—Consider the shock conditions along the rear shock in the form

$$[u_\alpha] = \frac{2(\rho u_n^2 - \gamma p)\nu_\alpha}{(\gamma + 1)\rho u_n}, \tag{1}$$

$$[p] = \frac{-2(\rho u_n^2 - \gamma p)}{\gamma + 1}, \tag{2}$$

$$[\rho] = \frac{-2\rho(\rho u_n^2 - \gamma p)}{2\gamma p + (\gamma - 1)\rho u_n^2}, \tag{3}$$

where the brackets [ ] have the usual significance and γ is the usual gas constant. The unit normal vector ν in (1) is directed as shown in figure 1. Also p and ρ stand for pressure and density immediately behind the rear shock and the quantity u_n denotes the normal velocity behind this shock.

Dividing both members of (2) by p we have

$$1 - \frac{p_2}{p} = \frac{-2\gamma}{\gamma + 1}\left(\frac{\rho u_n^2}{\gamma p} - 1\right),$$

where the subscript 2 is used to denote the value of a quantity immediately in front of the rear shock. Denoting by v the flow velocity immediately behind the shock let us put $c^2 = \gamma p/\rho$ and $m^2 = v^2/c^2$. Also $u_n = v \sin(\alpha - \bar{\omega})$, where α and $\bar{\omega}$ are the angles shown in figure 1. Then the above equation can be written

$$\frac{p_2}{p} = \frac{2\gamma}{\gamma + 1}[m^2 \sin^2(\alpha - \bar{\omega}) - 1] + 1. \tag{4}$$

Similarly we deduce from (3) that

$$\frac{\rho_2}{\rho} = \frac{2\gamma[m^2 \sin^2(\alpha - \bar{\omega}) - 1]}{(\gamma - 1)m^2 \sin^2(\alpha - \bar{\omega}) + 2} + 1, \tag{5}$$

and also by combining the two equations (1) we can obtain the following relation

$$\tan \omega = \frac{2[(m^2 - 1)\tan^2(\alpha - \bar{\omega}) - 1]}{[(\gamma - 1)m^2 + 2]\tan^3(\alpha - \bar{\omega}) + [(\gamma + 1)m^2 + 2]\tan(\alpha - \bar{\omega})} \tag{6}$$

between the angles α, $\bar{\omega}$ and the inclination ω of the stream line immediately in front of the rear shock.

3. Extension of the Pressure Equation.—Let $P'Q'$ be a curve ω = const. joining a point P' on the segment TR of the airfoil to a point Q' of the rear shock. We assume that the rear shock approaches the Mach direction of the free stream as the point Q' approaches infinity. Then the curves ω = const. establish a (1, 1) cor-

respondence $P' \leftrightarrow Q'$ between points P' on TR and points Q' on the rear shock such that $R \leftrightarrow R$ at the tail of the airfoil. Neglecting entropy changes the pressure p, which in general is a function $p(\omega, S)$ of the inclination ω of the stream lines and the entropy S, becomes a function $p(\omega)$ and hence the pressure as determined at Q' by the rear shock conditions will be approximately equal to the pressure at the corresponding point P'. Because of the above (1, 1) correspondence the pressure at all points of the segment TR will be obtained by this procedure.

We now assume that the flow behind the rear shock is identical with the flow in the free stream. Then if p_1 and ρ_1 are the pressure and density in the free stream we have $p = p_1$, $\rho = \rho_1$, and $m = M$ where M is the Mach Number of the free stream. Hence equation (4) becomes

$$\frac{p_2}{p_1} = \frac{2\gamma}{\gamma + 1} [M^2 \sin^2 (\alpha - \bar{\omega}) - 1] + 1. \tag{7}$$

The angles α and $\bar{\omega}$ are positive in equation (7) with $\alpha > \bar{\omega}$. Now under the above assumption the angle ω is obviously negative since the flow could not otherwise, after deflection by the shock through the angle $\bar{\omega}$, acquire the direction of the free stream for a shock line having the character of that shown in figure 1. It is therefore clear that $\omega = -\bar{\omega}$. Making this substitution in (6) we have

$$\tan (-\bar{\omega}) = \frac{2[(M^2 - 1) \tan^2 (\alpha - \bar{\omega}) - 1]}{[(\gamma - 1)M^2 + 2] \tan^3 (\alpha - \bar{\omega}) + [(\gamma + 1)M^2 + 2] \tan (\alpha - \bar{\omega})} \tag{8}$$

Now equations (7) and (8), in which the quantity $\alpha - \bar{\omega}$ can be thought of as a parameter, express the pressure ratio p_2/p_1 along the rear segment TR of the airfoil as a function of the inclination $-\bar{\omega}$ of this segment. That this formula for the pressure is an extension of the formula for the pressure over the segment VT, as derived from a consideration of the front shock conditions, is seen from the fact that the latter equation is obtained from the above equations (7) and (8) by replacing the parameter $\alpha - \bar{\omega}$ by α and $-\bar{\omega}$ in the right member of (8) by the inclination ω of the segment VT.

4. Entropy Change Across Rear Shock.—Denoting by N the quantity in the relation $p = N\rho\gamma$ the entropy S is defined by the equation $S = \log N/N_1$ where the subscript 1 refers to the free stream. By the entropy condition S cannot decrease in passage across the shock. Actually for compression shocks, which we assume in this discussion, the entropy will be greater at points immediately behind the shock than at contiguous points immediately in front of the shock.[2] Hence we must have $N/N_1 > 1$ at the point A in figure 1. But $N_1/N > 1$ at the point B on the same stream line as A (see figure 1) under the assumption that the flow behind the rear shock is the same as the flow in the free stream. Since the value of N is constant along stream lines and hence is the same in each of these inequalities we have thus arrived at a contradiction. In other words the flow behind the rear shock cannot strictly be identical with the free stream flow.

5. Assumption of Irrotational flow.—Let us now generalize the conditions in § 4 by making the assumption that the flow behind the rear shock is irrotational. Under this assumption it is well known that the quantity N will not merely be con-

stant along stream lines but constant throughout the entire flow. But as the point $Q' \to \infty$ along the rear shock the value of N will approach the value N_1 of this quantity in the free stream. Hence $N = N_1$ and we again have the contradiction obtained in § 4.

6. Natural Boundary Conditions.—Removing the above restrictions let us suppose that the flow behind the rear shock is rotational and thus has the same general character as the flow between the two shock lines in figure 1. Now it is natural to assume that as we approach infinity behind the rear shock in the direction of the free stream the flow quantities, i.e., the pressure, density, and velocity, will approach the values of these quantities in the free stream. But this means that $N \to N_1$ and since N is constant along stream lines it follows that N will have the constant value N_1 everywhere behind the rear shock. Hence the consideration in § 4 again applies and leads to a contradiction. We have commented briefly on this state of affairs in § 1.

* Research assistant under Navy Contract N6ONR-180, Task Order V, with Indiana University.

† Consultant, Naval Research Laboratory, Washington, D. C.

[1] See Brown, W. F., and Thomas, T. Y., "Limiting Behavior of Pressure Derivatives Behind Shocks in Supersonic Gas Flow," to appear in *J. Rat. Mech. Analysis*, 1954, §1. Also earlier papers by Thomas, T. Y., for example, "First Approximation of Pressure Distribution on Curved Profiles at Supersonic Speeds," these PROCEEDINGS, **35**, 617–627 (1949).

[2] See, for example, Thomas, T. Y., "The Fundamental Hydrodynamical Equations and Shock Conditions for Gases," *Math. Mag.*, **22**, 187 (1949). It can be inferred from the discussion here that $N = N_1$ implies $\rho_2 = \rho_1$ which contradicts the assumption that we have a compression shock.

GENETIC CHANGES IN AMERICAN POPULATIONS OF DROSOPHILA MELANOGASTER

BY P. T. IVES

DEPARTMENT OF BIOLOGY, AMHERST COLLEGE, AMHERST, MASS.

Communicated by Th. Dobzhansky, December 28, 1953

In earlier reports[1, 2] data were presented showing the frequency of genetically lethal second chromosomes in natural populations of *Drosophila melanogaster* in a few widely separated areas in the United States. The frequency of lethal chromosomes was so high in these areas, and the amount of allelism among the lethals was so low in the three areas in which samples of flies were taken directly from nature, that it seemed most probable that breeding units within populations of this species average considerably larger than in those of other species of Drosophila which had been reported at that time.

In the earlier studies with *D. melanogaster*, in all areas except Florida, samples were taken only in the early fall, when the natural population was close to peak size. Beginning in 1946 samples were taken from the local Massachusetts area at three times in the season: The first in June or July as soon as a large enough sample (about 100 flies minimum) could be collected within a week at one site to make the analysis worth while; the second about a month later; and the last collection as

late in the season as possible before the first killing frost in September or October. Beginning in 1950 a number of other areas were sampled in the early fall, and the Florida site was sampled in the spring of 1951. In most of these samples allelism tests were also completed with the strictly lethal chromosomes.

The only change in method from the earlier work consisted of the substitution of a new marker stock, known as *Sifter* to Drosophila geneticists, to facilitate making the test matings. Concurrent tests with *Bl/Cy* and *Sifter* in several different seasons showed no significant difference in lethal frequency (including semi-lethals) in the two methods of analysis.

Table 1 shows the frequency of lethals in different geographic areas, all collections from each area being summed.[3] Each collection site was in a rural or suburban area. None was within a large city area. It can be seen that the lethal frequency was higher in the three southernmost areas, Virginia, Texas, and Florida, than in the other more northerly localities. Lethals include all cases with less than 17% of homozygotes, as in the 1945 report. A comparison with the similar data of table 2 of that report shows that the lethal frequency is about 15% lower in these later

TABLE 1

FREQUENCY OF LETHALS IN DIFFERENT AREAS OF THE UNITED STATES

DATE	AREA	TESTS	LE	%
1947–1952	Amherst, Mass.	3163	1105	34.9
1950	New York State	527	170	32.3
1952	Cannonsburg, Pa.	117	33	28.2
1951	Wooster, Ohio	166	60	36.2
1952	Lincoln, Neb.	133	34	25.6
1951	Pullman, Wash.	138	54	39.1
1950–1952	Blacksburg, Va.	805	346	43.0
1952	Austin, Texas	98	41	41.8
1951	Winter Park, Fla.	131	67	51.1
Totals		5278	1910	36.2

samples from each of the three areas represented in both tables, Massachusetts, Florida, and Ohio. (In Ohio the two collection sites were in neighboring small city areas.) The other areas in the present study are also lower than would have been expected from the earlier study.

Tables 2 and 3 show in detail the change that has taken place in the two areas most critically sampled, Massachusetts and Florida. Items 1 and 2 of table 2 are from the 1945 report. Item 3 is from the 1947 report but has been corrected for the amount of allelism by removing those allelism tests involving lethal chromosomes derived from collected females. All samples in table 2 were collected in September or October, and the allelism data are for lethal chromosomes derived from collected males. In table 3 the 1940, 1942 data are from the 1945 report. In the 1951 Florida sample, females and males were separated soon enough after collection so that no increase in frequency of alleles was observed among lethal chromosomes derived from the females of the sample. In this case, therefore, the allelism data include tests of lethal chromosomes derived from both males and females.

Table 2 shows that apparently the major change in both lethal frequency and amount of allelism appeared first in 1947. The frequency dropped 10% from the

previous year and the amount of allelism approximately doubled compared to the average of the 1938–1946 period. The clearest statistical difference, however, is between the averages of the two groups of years, items 5 and 12. For the frequency of lethals, P is less than .0001, and for the amount of allelism it is approximately .0013, as determined by chi-square. The variation within each group of years is not significant, since the value of P is above .10 for each of the four comparisons.

Inspection of the data of table 2 shows, however, that there is actually considerable variation in lethal frequency within each group of years, the statistical significance of which may be obscured by the necessarily limited amount of data which could be collected each year. Particularly is this true in the 1947–1952 period where it looks as though the lethal frequency dropped secondarily in 1949 and came

TABLE 2

FREQUENCY OF LETHALS AND ALLELES IN DIFFERENT SEASONS IN SOUTH AMHERST, MASS.

ITEM	YEAR	TESTS	LETHALS	%	CROSSTESTS	ALLELES	%
1	1938	151	68	45.0	1176	5	0.43
2	1941	108	64	59.3	None		
3	1945	190	87	45.8	666	3	0.45
4	1946	79	39	49.3	465	1	0.22
5	Total 1–4	528	258	48.8	2307	9	0.39
6	1947	257	101	39.3	1039	9	0.87
7	1948	162	60	37.0	1219	11	0.90
8	1949	198	60	30.3	1223	11	0.90
9	1950	163	52	31.9	528	4	0.76
10	1951	182	68	37.4	1377	17	1.24
11	1952	190	72	37.9	820	5	0.61
12	Total 6–11	1152	413	35.9	6206	57	0.92

TABLE 3

THE FREQUENCY OF LETHALS AND ALLELES AT DIFFERENT TIMES IN WINTER PARK, FLORIDA

YEARS	TESTS	LETHALS	%	CROSSTESTS	ALLELES	%
1940, 1942	337	220	65.3	2281	10	0.44
1951	131	67	51.1	1326	5	0.38

back up to its 1947–1948 level in 1951–1952. The value of P for the six years as a group is .3; for the comparison of the three pairs of years (paired chronologically to give two degrees of freedom) it is .07; and for the comparison of 1949–1950, first with 1947–1948, and then with 1951–1952 it is .04 and .06, respectively. The P values are large enough so that the possibility of this variation being purely random cannot be considered unlikely. At the same time they are small enough to render more acceptable the alternative interpretation that there was a U-shaped decrease and increase in lethal frequency in the 1947–1952 period. There is no suggestion of any related inverse change in the amount of allelism observed in this period, perhaps because the data are not large enough to demonstrate comparatively small changes in this category. It is probably only coincidental that this apparent cycle of change occurred at exactly two-year intervals in this period of time.

The data of table 3 show that the drop in lethal frequency (P is .0007) in the

Florida areas was not accompanied by a rise in the observed amount of allelism among lethals. In 1951 the lethal frequency and the amount of allelism in the Florida area each resembles closely the corresponding average in the Massachusetts area in the 1938–1946 period.

Some additional information on the Massachusetts population is given in table 4. These data show for the 1947–1952 period the average lethal frequency in the first, second, and last collections and the average amount of allelism in the first and last collections. The data are for five of the six seasons; in 1949 it was possible to make collections only in August and October. Inspection of the data suggests that in the time between the first and second collections, when the population was expanding rapidly, there was an increase in lethal frequency, but that in the rest of the season there was no change. It also suggests that there was the same amount of allelism among lethals at both phases of population development, expansion and peak.

In five of the six seasons in which the first collection came before August (including 1946) there was an increase in lethal frequency in the second collection; 1952 was the exceptional year. The increase was not statistically significant in any one season. Considering only the chi-square P value of .08, the average increase in table 4 is on the borderline of significance statistically. If the increase is real (the

TABLE 4

THE 1947–1952 AVERAGE INTRA-SEASONAL FREQUENCIES OF LETHALS AND ALLELES IN SOUTH AMHERST, MASS.

SAMPLE	TESTS	LETHALS	%	CROSSTESTS	ALLELES	%
First	726	233	32.1	2628	29	1.10
Second	550	202	36.7			
Last	954	353	37.0	4983	46	0.92

author feels that more data are needed to determine that fact) it can probably be satisfactorily accounted for by two or three generations of accumulation of new mutations.[4] The increase is not as large as the average increase in Russian populations.[5] Goldschmidt has reported no increase during the expansion phase of Jerusalem populations.[6] If new mutations are the cause in the present case, one must assume an earlier elimination of lethals in the comparatively dormant season, or in the earlier part of the breeding season than is represented in the first collections, due either to inbreeding or selection. There are no data on the amount of allelism at actual breeding sites in the local area.

The complete picture is not given with respect to allelism in the simple averages of table 4. In some seasons it was possible to test for allelism within and between three or more collections per season. The details of this study will appear elsewhere. They will show that each season probably differs from the others in the nature of the breeding units which send flies to the collection site. The picture is much more dynamic and complex than the averages given here suggest. The present data do show, nevertheless, that in this group of seasons the average amount of allelism among the lethal chromosomes was approximately the same in a potential gene pool at the time when the population was in its stage of rapid expansion as it was a little later on when the peak of population size had been reached.

It is apparent from the data of tables 1, 2, and 3 that there was a significant

change in the genetic structure of natural populations of *D. melanogaster* in America during the past decade. One simple interpretation is that there has been a decrease in population size, in the number of pairs of individuals which constitute the average breeding unit. This would cause an increase in the amount of inbreeding, which would in turn be reflected in an increase in allelism and a decrease in the frequency of lethal chromosomes. That description fits the data of the local area. Another possibility is an increased selection against lethal heterozygotes, which would lower the lethal frequency without necessarily changing the amount of allelism noticeably, such as appears to be the case in the Florida data of table 3. There is no conclusive evidence concerning these interpretations.

It is possible that in the most southerly regions *D. melanogaster* is suffering from competition with its closely related species, *D. simulans*. With collection conditions as alike as it was possible to make them, the sample of 1942 from Florida contained an estimated 15% of simulans in about 300 males, while in the 1951 sample of 990 flies there was 86% simulans, with no difference in percentage of simulans in males and females. The small sample from Austin, Texas, in 1952 had 40% simulans in the 116 males of the collection. *D. simulans* was very infrequent in the collections from other areas. It has appeared locally only in one or two males per season in late summer collections.

The causative agent for the presumed decrease in size in the northerly populations of *D. melanogaster* during the 1940's is not known. The change in climate to a warmer and drier summer, generally, has not been sufficient locally to interfere noticeably with the presumed food source of the species, decaying fruits, vegetables, and garbage, at least so far as can be seen by casual observation. The introduction of new fungicide and insecticide sprays in orchards and buildings may be a factor. So far as can be determined there was no change in spraying program in 1947 as compared to 1946 but several new sprays were introduced in subsequent seasons. These might have contributed to the apparent population drop in 1949 and 1950. Whatever the cause the species seems to have adapted itself to the situation and may have improved its position in 1951 and 1952.

It seems obvious that experiments with large continuous cage populations may suggest the nature of the influence of certain environmental agents on the genetic factors of natural populations, beginning with some of the genetic factors already under investigation.[7–10] At the same time it should perhaps be emphasized that in a natural population the combination of important variables and effective agents is probably so complex that only very general applications can be made of the findings from studies of artificial cage populations. It is not possible to "experiment" with such variables and agents and at the same time maintain truly natural populations. Much remains to be done of value, however, at the descriptive genetic level in the study of natural populations of *Drosophila melanogaster*.

[1] Ives, P. T., *Genetics*, **30**, 167–196 (1945).

[2] Ives, P. T., *Evolution*, **1**, 42–47 (1947).

[3] The author is indebted to the following persons for collecting flies: Dr. Mary Alexander, Dr. R. P. Levine, Dr. M. Levitan, Dr. D. D. Miller, Dr. R. Moree, Mrs. H. H. Plough, Mr. M. Seiger, and Dr. W. P. Spencer. He is also grateful to Dr. B. Wallace for the use of his unpublished data from three samples of 449 chromosomes which are included in the New York State sample after the individual tests had been reclassified by the standard for lethals used in the other samples of the present study.

[4] Ives, P. T., *Evolution*, **4**, 236–252 (1950).
[5] Dubinin, N. P., *Genetics*, **31**, 21–38 (1946).
[6] Goldschmidt, Elizabeth, *Drosophila Information Service*, **26**, 102–103 (1952).
[7] Wallace, B., *Evolution*, **4**, 172–174 (1950).
[8] Wallace, B., and King, J. C., these PROCEEDINGS, **38**, 706–715 (1952).
[9] Wallace, B., and Madden, Carol, *Genetics*, **38**, 456–470 (1953).
[10] Levine, R. P., and Ives, P. T., these PROCEEDINGS, **39**, 817–823 (1953).

PSEUDO-ALLELISM AT THE VERMILION LOCUS IN DROSOPHILA MELANOGASTER

BY M. M. GREEN

DEPARTMENT OF GENETICS, UNIVERSITY OF CALIFORNIA, DAVIS, CALIFORNIA

Communicated by R. E. Clausen, December 10, 1953

A number of cases have been reported in Drosophila[1–6] and Aspergillus,[7, 8] and apparent cases in Neurospora[9, 10] and cotton,[11] in which recombination occurs between mutants which otherwise classify as allelic by the usual phenotypic and genetic criteria. Such mutants have been designated as pseudo-alleles. In addition to recombination, pseudo-alleles manifest a differential phenotype depending upon whether the mutants are compounded in coupling or in repulsion. Thus in a case of pseudo-alleles which can be designated as m_1 and m_2, both of which are recessive to wild-type, the phenotype of individuals of the genotype $m_1 + / + m_2$ is mutant, while that of individuals $m_1\ m_2 / + +$ is wild-type. Alternative explanations have been proposed to account for this situation. On the one hand, it has been reasoned that pseudo-alleles represent separate genes whose loci and functions are distinct. The phenotypic differences alluded to are then interpreted as being the consequence of position effect.[1, 2, 4] Alternatively, it has been argued that pseudo-alleles are components of a larger "physiological" gene, whose functions are indistinguishable but whose component parts are separable by crossing-over.[8, 12] The phenotypic differences associated with pseudo-alleles are therefore those expected according to the usual rules of dominance.

These alternative interpretations imply conflicting definitions of the gene. On the one hand the gene as defined in terms of linkage relations, mechanism of action and mutability is one and the same. On the other the limits determined for the gene in terms of linkage do not necessarily coincide with the limits of the gene in terms of function.

It is the purpose of this report to discuss a case of pseudo-allelism at the vermilion eye color (v) locus in *D. melanogaster* and to determine what bearing it has on the interpretation of pseudo-allelism and the gene.

Phenotypic Differentiation of v Mutants.—In an earlier discussion of several sex-linked, recessive v mutants of *D. melanogaster*, it was noted that while they are essentially phenotypically indistinguishable, it is possible to separate them into two classes on the basis of their phenotypes in the presence of a non-allelic suppressor mutant.[13] Certain of the v mutants (collectively designated as v^s) are suppressed, while others (collectively designated as v^u) are unsuppressed. It was

also noted that v^s and v^u mutants react in identical fashion in gynandromorphs and when developing larvae are supplied either formyl-kynurenine or kynurenine.[13] Under the latter conditions both types of v mutants synthesize brown eye pigment.

Additional data are now available which serve for further comparison of the v mutants. It has been known that v^s mutants when subjected to partial larval "starvation" synthesize brown eye pigment.[14] Tests of the effects of larval "starvation" on the eye color of v^u mutants have been performed and demonstrate that they fail to synthesize brown eye pigment under these conditions,[15] thereby presenting a second means of distinguishing between v mutants. It has also been reported that v^s mutants raised in an optimal nutritional regime accumulate non-protein tryptophane.[16] A comparison of the non-protein tryptophane contents of v^s and v^u mutants cultured under such conditions shows that they accumulate like amounts of non-protein tryptophane, and it is not possible to distinguish between them on this basis.[15] Table 1 comprises a summary of the several phenotypic and biochemical comparisons which have been made with the v mutants. It may be noted that only the suppressor and larval "starvation" serve to distinguish among the v mutants.

TABLE 1

COMPARATIVE PHENOTYPIC RESPONSES OF THE TWO CLASSES OF v MUTANTS

v MUTANT	SUPPRESSOR	LARVAL "STARVATION"	FORMYL-KYNURENINE	KYNURENINE	TRYPTOPHANE ACCUMULATION
v^s	+	+	+	+	yes
v^u	−	−	+	+	yes

+ = brown pigment formed; − = no brown pigment formed.

TABLE 2

TESTS OF v MUTANTS TO SUPPRESSOR, su^2-s

ORIGIN OF MUTATION	SUPPRESSED	UNSUPPRESSED	N
Spontaneous	3	3	6
X-ray and gamma	0	15	15
Ultra-violet	0	1	1

TABLE 3

CROSSING-OVER BETWEEN v^1 AND v^{36f} (EXPERIMENTS 1–3, ♀ ♀ WITH FREE X CHROMOSOMES; EXPERIMENTS 4 AND 5 ♀ ♀ WITH ATTACHED-X CHROMOSOMES)

	GENOTYPE OF ♀ ♀			
	X chromosomes;	Autosomes*	v^+ ♂ ♀	N
1.	$v^{36f}/ras^2\ v^1\ m\ f$;	$Cy/+$; $In(3)CXF$, $D/+$	1 ♂ = $v^+\ m\ f$	34,556
2.	$v^{36f}/ras^2\ v^1\ m\ f$;	$Cy/+$; Sb, $In(3R)C/+$	1 ♂ = $v^+\ m\ f$	33,600
3.	$v^{36f}/ras^2\ v^1\ m\ f$;	$Cy/+$; $In(3)CXF$, D/Sb, $In(3R)C$	0	37,448
4.	$ras^2\ v^1\ f/v^{36f}\ m\ f$;	$Cy/+$; $Ubx/+$	1 ♀ = $v^+\ f$	45,651
5.	$ras^2\ v^1\ f/v^{36f}\ m\ f$;	$Cy/+$; $In(3)T$, $Me/+$	3 ♀ ♀ = $v^+\ f$	23,951

* See Bridges and Brehme[20] and Lewis[5] for description of autosomal inversions.

Crossing-over Between v Mutants.—It was suggested earlier in view of the separability of v mutants on the basis of the suppressor that the v locus is pseudo-allelic.[13] This possibility has been tested, using the mutants v^1 (suppressed) and v^{36f} (unsuppressed). In these experiments the mutants raspberry-2 eye color (ras^2), 0.2 unit to the left, and miniature wing (m) 3.1 units to the right of the v locus, were used as crossing-over markers. Females of the genotype $ras^2\ v^1\ m/v^{36f}$ and heterozygous for several autosomal inversions were constituted and crossed to $ras^2\ v^1\ m$

♂♂. The results of these crosses are listed in table 3. It may be noted that two v^+ m ♂♂ were recovered. The occurrence of v^+ individuals associated with crossing-over is indicative of pseudo-allelism. This may be best accounted for here by assuming that v^1 lies just to the left of v^{36f} such that the genotype of the parental ♀♀ was $ras^2\ v^1\ +m/+\ +\ v^{36f}\ +$. A single crossover occurring between v^1 and v^{36f} would yield $+\ +\ +\ m$ or $ras^2\ v^1\ v^{36f}$ individuals.

It may be noted that on the basis of pseudo-allelism, in addition to the v^+ m individuals recovered from ♀♀ $ras^2\ v^1\ m/v^{36f}$ the complementary crossover type, viz., $ras^2\ v^1\ v^{36f}$ should occur. There was no evidence for the occurrence of this type

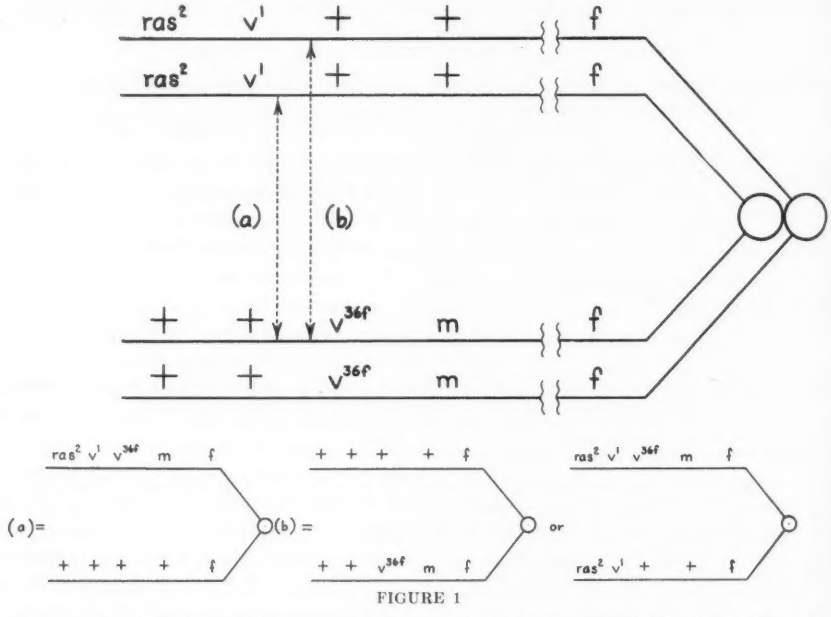


FIGURE 1

Scheme for detecting crossing-over between pseudo-alleles v^1 and v^{36f} in attached-X ♀♀. The upper diagram represents the genotype of the attached-X ♀♀. The reciprocal crossover is designated (*a*), the non-reciprocal (*b*), and the corresponding crossover products shown below are designated (*a*) and (*b*).

in the initial crosses. In fact, it appears highly unlikely that individuals of the genotype $ras^2\ v^1\ v^{36f}$ could be differentiated phenotypically from ♀♀ $ras^2\ v^1$ or $ras^2\ v^{36f}$. Since both the v^1 and v^{36f} mutants are incapable of utilizing tryptophane, it appears reasonable to conclude that the mutants coupled together on the same X chromosome would also fail to utilize tryptophane. Accordingly, it would be expected that the phenotype of individuals homozygous for $v^1\ v^{36f}$ would be inseparable from that of individuals homozygous v^1 or v^{36f}.

There does exist at least one method which permits the recovery from ♀♀ v^1/v^{36f} of both products of crossing-over between the pseudo-alleles. This involves crossing-over in attached-X ♀♀. This technique was first utilized for recovering

simultaneously both products of unequal crossing-over in Bar eye in *D. melanogaster*,[17] and has more recently been used to establish pseudo-allelism at the white locus in *D. melanogaster*.[5] Accordingly attached-X ♀ ♀ of the genotype *ras*^{2} *v*^{1} + *f*/+ + *v*^{36f} *m f* were constituted, autosomal inversions introduced, and these ♀ ♀ were then crossed to *v f* ♂ ♂. The ♀ progeny of these crosses were examined and *v*$^+$ *f* ♀ ♀ were sought. In figure 1, the genotype of the attached-X ♀ ♀ is illustrated; in addition, the products of crossing-over between the *v* pseudo-alleles are listed. As indicated in figure 1, three products of crossing-over between the pseudo-alleles can occur, but only the two producing *v*$^+$ phenotypes are detectable. The *v*$^+$ phenotypes will occur as a result of either a reciprocal or a non-reciprocal crossover. It may be noted that *v*$^+$ ♀ ♀ occurring as a result of a reciprocal crossover between the pseudo-alleles will possess both chromatids involved in the crossover, i.e., will have one X chromosome carrying the wild-type alleles of the *v* pseudo-alleles while the homologous X will carry the two mutants pseudo-alleles. These ♀ ♀ will be of the genotype indicated for type (*a*) in figure 1 and will be phenotypically *v*$^+$ *f*. That such is the case can be determined by a progeny test. If the *v*$^+$ *f* ♀ is the result of a reciprocal crossover, there should occur among her attached-X ♀ progeny individuals of the phenotype *ras*^{2} *v m f*. These, the result of a subsequent non-reciprocal crossover occurring to the right of *m* in the *v*$^+$ *f* ♀, represent the individuals homozygous for both pseudo-alleles and each X chromosome is genotypically *ras*^{2} *v*^{1} *v*^{36f} *m f*.

The *v*$^+$ *f* ♀ ♀ occurring as a result of non-reciprocal crossover between the pseudo-allelic loci will possess one crossover chromatid and one non-crossover chromatid. Their genotype corresponds to (*b*) of figure 1 and also will be phenotypically *v*$^+$ *f*. On progeny testing, these ♀ ♀ should produce among the attached-X ♀ progeny the phenotype *v*^{36f} *m f*. This occurs as a result of a non-reciprocal crossover occurring to the right of *m* in the *v*$^+$ *f* ♀.

In the initial tests made as listed in 4 of table 3, one *v*$^+$ *f* attached-X ♀ was recovered. On progeny testing, this ♀ produced a total of 113 attached-X ♀ progeny, of which 105 were phenotypically *v*$^+$ *f* and 8 *ras*^{2} *v m f*. Since the eye color of *ras*^{2} *v*^{1} *v*^{36f} *m f* ♀ ♀ is inseparable from that of ♀ ♀ *ras*^{2} *v*^{1} and since the ♀ ♀ carry the markers expected for a reciprocal crossover, it may be concluded that these attached-X ♀ ♀ are genotypically homozygous *ras*^{2} *v*^{1} *v*^{36f} *m f*. The attached-X's of ♀ ♀ *ras*^{2} *v*^{1} *v*^{36f} *m f* were detached and tested to the suppressor. As expected, the suppressor failed to alter the eye color. Thus it can be concluded that the crossover carrying the two pseudo-alleles has been recovered.

Three additional cases of crossing-over between the *v* pseudo-alleles in attached-X ♀ ♀ have been recovered in an experiment where a different combination of autosomal inversions was used. These are listed in 5 of table 3.

Cytology of the v Locus.—In a number of cases of pseudo-allelism in *D. melanogaster*, evidence has been submitted which points strongly to the association of the pseudo-alleles with a doublet as observed in the salivary gland chromosomes.[4] To determine whether such an association exists in the case of the *v* pseudo-alleles four independent *v* deficiencies were examined cytologically. While the limits of the deficiencies varied, all had one feature in common, viz., the absence of the doublet 10A 1–2 as listed in Bridges' revised salivary gland X chromosome map.[18] The association of the *v* locus with a cytologically visible doublet supports the interpreta-

tion of pseudo-allelism, but it cannot be concluded with absolute certainty that v^1 is localized in the left half (10 A1) and v^{36f} in the right half (10 A2) of the doublet.

Mutation to v^s and v^u.—In the course of this study a total of 22 independently occurring v mutations have been assembled and tested to the suppressor. These mutants, a number of which were contributed by other investigators, all arose from v^+ and include both spontaneous and induced mutants. In table 2 the results of the tests of the mutants with the suppressor together with their mode of origin are summarized. It is significant to note that whereas among the spontaneous mutants both v^s and v^u types occur, only v^u mutants have been found in the induced group. Or phrased slightly differently, while v^s mutants occur spontaneously, they have thus far failed to occur among the induced v mutants. Does this observation mean that there is a difference in induced mutability between the pseudo-allelic v loci? This question cannot be conclusively answered at this time for a number of reasons. The data on crossing-over between the pseudo-alleles have been derived from spontaneously occurring mutants. Thus far crossing-over between a v^s mutant and an x-ray induced v^u mutant has not been observed although admittedly extensive trials have not been made. Two x-ray v^u mutants were tested with v^1 in attached-X ♀♀, paralleling experiment 4 of table 3. In each case no v^+ ♀♀ were found among 40,000 progeny. These negative results are not conclusive. If the observed crossing-over between v^1 and v^{36f} is a reasonable approximation of the frequency of crossing-over between the pseudo-alleles (see table 3), the negative results noted can readily be accounted for on the basis of chance. The relatively low frequency of crossing-over between v^1 and v^{36f} has for the time being discouraged large scale testing for crossing-over between other v mutants.

If it is assumed that the suppressor does distinguish between the pseudo-allelic loci, it is possible to interpret the origin of induced v^u mutants in one of three ways: (1) All induced mutants are "point" mutations allelic to v^{36f}. (2) All induced mutants involve both loci either as small rearrangements or "point" mutations at both loci. (3) Induced mutants include both types (1) and (2).

The available genetic and cytological information indicates that none of the induced v^u mutants is associated with gross chromosomal rearrangement. While it cannot be stated conclusively at this time whether induced v^u mutants involve both pseudo-allelic loci, there is evidence available from studies of lozenge (*lz*) pseudo-alleles[2, 19] that *lz* mutants rarely if ever involve two loci simultaneously. Analysis of eight independent x-ray induced *lz* mutants demonstrates that each can be assigned, on the basis of crossing-over, to one of the three *lz* loci. If the induced v mutants are homologous to the induced *lz* mutants in that only one locus mutates at a time, and if the suppressor does distinguish between the pseudo-allelic loci, then the results noted demonstrate differential mutability of the v pseudo-allales. It is pertinent to point out here that the observations on induced mutation are not peculiar to the v pseudo-alleles. In a study now in progress[19] on the forked (*f*) locus in *D. melanogaster*, it has been shown that it is possible to distinguish between *f* mutants by using a non-allelic suppressor mutant; pseudo-allelism is indicated for the *f* locus since crossing-over occurs between suppressed and unsuppressed *f* mutants; and all induced *f* mutants studied are unsuppressed. The analysis of the *f* locus will be discussed shortly.

Discussion.—The foregoing account of the genetics and cytology of mutants at

the *v* locus in *D. melanogaster* has established the existence of two pseudo-allelic *v* loci. It is of importance to reconsider here the bearing this case of pseudo-allelism has upon the general interpretation of pseudo-allelism. As emphasized at the outset, the specific question which the phenomenon of pseudo-allelism poses may be stated as follows: do pseudo-alleles represent closely linked genes with distinctive functions or do they represent an integrated functional unit of the chromosome whose component parts can be separated by means of crossing-over? At first glance, it would appear that the latter interpretation more closely fits the *v* pseudo-alleles. The phenotypic effects of the several *v* mutants alluded to previously indicate that all *v* mutants are incapable of utilizing tryptophane, the initial step in the biosynthesis of brown eye pigment. This is borne out by the observations that all *v* mutants accumulate non-protein tryptophane and respond similarly by synthesizing brown eye pigment when supplied either formyl-kynurenine or kynurenine. Thus, it would appear that *v* mutants are functionally inseparable. There are, however, two observations which militate against an interpretation of functional identity among the several *v* mutants studied. These are the formation of brown eye pigment by some but not all mutants, either in the presence of a non-allelic suppressor or when subjected to partial larval "starvation." It should be added here that there is a slight but consistent phenotypic difference which the experienced observer can use to distinguish between *v* mutants. The phenotypes of v^s and v^u individuals can be distinguished when in addition they are homozygous for the autosomal eye color mutant brown (*bw*). Thus while the eye color of v^u; *bw* flies is essentially white that of v^s; *bw* is not and a discernible quantity of brown eye pigment is formed.

The role of the suppressor mutant has significance for the interpretation and meaning of pseudo-allelism. As noted previously, the suppressor mutant su^2-*s* not only suppresses some *v* mutants but also the sex-linked recessive body color mutant sable (*s*). A number of alleles of su^2-*s* have been reported;[20] one has been described as suppressing not only *v* but the recessive, second chromosome eye color mutant purple (*pr*), while another has been described as suppressing the recessive, second chromosome mutant speck (*sp*). Accordingly, appropriate tests were made to determine whether su^2-*s* suppresses *pr* and *sp*. The results obtained demonstrated that su^2-*s* suppresses *pr* and *sp* in addition to *v* and *s*. An allele of su^2-*s*, su^9-*v*, also was found to suppress *pr* and *sp*.

These observations on the action of the suppressor support the contention that the v^s and v^u mutants have separate, distinctive actions. It is pertinent to point out here that *pr* eye color mutant is autonomous,[21] in contrast to both the v^s and v^u mutants which are non-autonomous in development.[13, 21] In addition *pr* flies do not accumulate non-protein tryptophane.[15] Thus the suppressor acts to suppress other mutants, but fails to suppress v^u mutants which superficially appear to be closely related in their mode of action to the v^s mutants. It therefore appears highly unlikely that the v^s and v^u mutants can be identical in their modes of action and yet be separable in their reactions in the presence of the suppressor. Since the mutants v^1 and v^{36f} have been separated by crossing-over, by the suppressor, and by larval "starvation," the conclusion can be drawn that these mutants constitute different genes with distinctive loci and functions. These findings for the *v* pseudo-alleles are in accord with those noted for the bithorax-bithoraxoid pseudo-alleles[4] which can be

separated not only by crossing-over, but in addition on the basis of their distinctive phenotypic effects. In addition, the case of the inositolless mutants in Neurospora, which can be distinguished by means of a suppressor and which apparently cross over, parallels the results reported here for the *v* pseudo-alleles.[9, 22]

The question of the basic difference in the mode of action of the mutants v^1 and v^{36f} cannot be satisfactorily answered at this time. It has been noted that both mutants are blocked biochemically at the initial step in the biosynthesis of brown eye pigment and thus both accumulate non-protein tryptophane. The investigation of the genetic control of tryptophane desmolase in Neurospora is a case in point.[23] Here, it has been ably demonstrated that two allelic mutants, separable by an independent suppressor, both influence the same enzyme activity. By analogy, the case of the *v* pseudo-alleles may also represent a case where the same enzyme system is altered by different mutant genes in different ways. The fact that v^s flies can synthesize brown eye pigment in the presence of the suppressor or when subjected to larval "starvation" suggests that they may, in fact, possess the necessary enzyme system which catalyzes the oxidation of tryptophane but that the particular enzyme(s) is either inactive or is inactivated by the v^s gene or its products. The suppressor or larval "starvation" thus serve to activate the enzyme(s) or to remove its inhibitors. In the case of the v^u mutants it would appear that a thus far irreversible alteration in the enzyme system has occurred such that it is incapable of catalyzing the oxidation of tryptophane. It must be emphasized that the foregoing interpretations of the mechanism of action of the v^s and v^u mutants are tentative. While there is no *a priori* reason to believe that the same enzyme system cannot be altered by different genes and therefore in different ways, there is as yet no information available which permits a critical discrimination among the possible mechanisms which might be operative here.

What interpretation can be submitted to explain the differential phenotype observed when the pseudo-alleles are in coupling versus repulsion? Since all *v* mutants are recessive to wild-type it readily follows that individuals of the coupling genotype $v^1\ v^{36f}/+\ +$ should be wild-type in phenotype. However, the fact that individuals of the repulsion genotype $v^1\ +/+\ v^{36f}$ are vermilion in phenotype clearly fails to follow the expectation based on dominance of the wild-type alleles. Thus any explanation of phenotype based on the assumption of simple dominance and recessivity fails. Rather it appears possible to interpret the resultant phenotypes in terms of the primary action of the mutant genes themselves along the lines previously discussed by Lewis.[4] It is assumed here that the primary action of both v^1 and v^{36f} mutants involves a block at tryptophane, each accomplishing the block in a different way. The interpretation of the phenotypes of the mutants in coupling and repulsion must include the assumption that the primary gene-controlled reactions are here confined to the immediate vicinity of the gene. Moreover, it must be assumed that the immediate reactions governed by the genes on one chromosome are independent of those on the homologous chromosome. Thus in the case of individuals of the genotype $v^1\ v^{36f}/+\ +$, the utilization of tryptophane is blocked by the *v* mutants on the one chromosome, but this block does not impede the action of the + alleles on the homologous chromosomes in utilizing tryptophane and thus brown eye pigment is synthesized. However in the case of the individuals of the genotype $v^1\ +/+\ v^{36f}$, each mutant pseudo-allele acting independently blocks the

utilization of tryptophane and thus prevents the biosynthesis of brown eye pigment. Thus while the wild-type alleles of each pseudo-allele are present, their effects are rendered ineffective by the action of the adjacent *v* mutants and the resultant phenotype is vermilion rather than wild-type.

Summary.—A phenotypic comparison between two types of *v* mutants in *D. melanogaster* is presented.

Pseudo-allelism at the *v* locus is indicated since crossing-over between the mutants v^1 and v^{36f} was obtained.

Cytologically the *v* loci appear to be associated with the doublet 10 A1-2 in the salivary gland chromosome.

The bearing of these results on the interpretation of pseudo-allelism and the nature of the genes involved is discussed.

[1] Lewis, E. B., *Genetics*, **30**, 137–166 (1945).
[2] Green, M. M., and Green, K. C., PROC. NATL. ACAD. SCI., **35**, 586–591 (1949).
[3] Ives, P. T., and Noyes, D. T., *Anat. Rec.*, **111**, 565 (1951).
[4] Lewis, E. B., *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 159–174 (1951).
[5] Lewis, E. B., PROC. NATL. ACAD. SCI., **38**, 953–961 (1952).
[6] MacKendrick, M. E., and Pontecorvo, G., *Experientia*, **8**, 309 (1952).
[7] Roper, J. A., *Nature*, **166**, 956 (1950).
[8] Pontecorvo, G., *Advances in Genetics*, **5**, 141–238 (1953).
[9] Giles, N. H., *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 283–313 (1951).
[10] Bonner, D. M., *Ibid.*, **16**, 143–157 (1951).
[11] Stephens, S. G., *Ibid.*, **16**, 131–141 (1951).
[12] Goldschmidt, R. B., PROC. NATL. ACAD. SCI., **36**, 365–368 (1950); *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 1–11 (1951).
[13] Green, M. M., PROC. NATL. ACAD. SCI., **38**, 300–305 (1952).
[14] Tatum, E. L., and Beadle, G. W., *Biol. Bull.*, **77**, 415–422 (1939).
[15] Shapard, P. B., unpublished data.
[16] Green, M. M., *Genetics*, **34**, 564–572 (1949).
[17] Morgan, L. V., PROC. NATL. ACAD. SCI., **17**, 270–272 (1931).
[18] Bridges, C. B., *J. Heredity*, **29**, 11–13 (1938).
[19] Green, M. M., unpublished data.
[20] Bridges, C. B., and Brehme, K. S., *Publ. Carnegie Inst. Wash.*, **552**, 1–257 (1944).
[21] Beadle, G. W., and Ephrussi, B., *Genetics*, **21**, 225–247 (1936).
[22] Giles, N. H., and Partridge, C. W. H., PROC. NATL. ACAD. SCI., **39**, 479–488 (1953).
[23] Yanofsky, C., *Ibid.*, **38**, 215–226 (1952).

---

*NUMERICAL PREDICTION OF CYCLOGENESIS**

BY J. G. CHARNEY

THE INSTITUTE FOR ADVANCED STUDY

Communicated by J. von Neumann, November 16, 1953

1. Introduction.—In a recent article[1] the writer and N. A. Phillips devised a sequence E_n of mathematical models of the atmosphere capable of predicting the geopotentials of n suitably selected isobaric surfaces, and then presented the results of a series of consecutive 24-hour numerical integrations that had been carried out with the models E_1 and E_2 for an intensely cyclogenetic situation.

The predictions were moderately accurate during the period immediately preceding the storm but deteriorated markedly in accuracy after its onset. Although a rudimentary type of development was observed to take place in the second model, in neither case could it be said that the cyclogenesis had been predicted.

These results were not unanticipated. The model E_1 was a single layer autobarotropic atmosphere in quasi-hydrostatic and quasi-geostrophic equilibrium. Such an atmosphere contains no effective source of kinetic energy and permits no individual change in vorticity. The model E_2 corresponded closely to an atmosphere consisting of two superimposed autobarotropic layers. An atmosphere of this type does permit potential-kinetic energy conversions as well as individual vorticity changes but does not permit horizontal variations of static stability and consequently imposes a severe artificial constraint upon the motion. For example, it cannot simulate the undoubtedly important effects of the low-level thermal asymmetries associated with a front.

At the same time it was not obvious that the above-mentioned shortcomings had been solely responsible for the failure to predict the cyclogenesis. Defects of a more fundamental nature might have been inherent in the postulates underlying the entire sequence of models. The fault might have resided in the quasi-geostrophic approximation, or possibly in the failure to take into account non-isentropic changes. It was even conceivable, had the view been held that the cyclogenesis was an instability phenomenon of the conventional type, that the error might have been due to the practical indeterminacy of the initial perturbation.

Since the three-level model E_3 is relatively free of the first class of defects but not of the second, numerical integrations with this model for the same periods could be expected at least to localize the sources of error. Such integrations were accordingly programmed and carried out on the Institute for Advanced Study (IAS) computer. The integrations showed conclusively that the errors had been of the former variety, for in fact the initiation and subsequent development of the storm were predicted. As this was perhaps the first instance of a purely mathematical prediction of cyclogenesis, it is believed that the presentation of the results and a discussion of their implications will be of general interest.

2. Description of Model.—The basic premise used in the construction of the models E_n is that the motion of the atmosphere departs only slightly from solid rotation with the earth. If the horizontal scale of the motion is sufficiently large, it then follows that the flow will be in quasi-hydrostatic and quasi-geostrophic equilibrium, i.e.,

$$\frac{\partial \phi}{\partial p} + \alpha \cong 0, \tag{1}$$

and

$$f\vec{v} \times \vec{k} + \nabla\phi \cong 0. \tag{2}$$

Here the pressure p is the vertical coordinate, ϕ is the geopotential, α the specific volume, $\vec{v}$ the horizontal relative velocity component, $\vec{k}$ the upwardly pointing vertical unit vector, ∇ the horizontal gradient operator, $f = 2\Omega \sin \varphi$, where Ω is the earth's angular velocity and φ is the latitude. If it is also assumed that the

flow is isentropic and non-viscous, it may be shown that the motion is governed entirely by the scalar conservation equations for entropy and potential vorticity q.[2] From these the equations for the sequence of models E_n are derived by a series of transformations and finite-difference approximations. If the earth is mapped conformally onto a plane with the cartesian coordinates x and y, and m is the scale factor, the equations take the form

$$\left(\frac{\partial}{\partial t} + \vec{v}_k \cdot \nabla\right) q_k = \frac{\partial q_k}{\partial t} - \frac{\partial(m^2 f^{-1} q_k + f, \phi_k)}{\partial(x, y)} = 0, \tag{3}$$

$$q_k = \nabla^2 \phi_k + \sum_{r=1}^{n} A_{rk} \phi_r \qquad (k = 1, 2, \ldots, n), \tag{4}$$

in which ϕ_k is the sole dependent variable and the matrix (A_{rk}) is constant and negative definite.

The boundary conditions for a region R in the plane bounded by the simple closed curve C are that the ϕ's must be specified as functions of time on C and the q's at those and only those points of C where fluid is entering R. Initially ϕ must be specified everywhere within and on the boundary of R.

The indices k refer to the isobaric surfaces

$$p_k = \frac{k - {}^1/_2}{n} p_0,$$

where p_0 is the mean surface pressure. Assuming $p_0 = 1000$ mb we get for $n = 3$: $p_1 = 167$ mb, $p_2 = 500$ mb, and $p_3 = 833$ mb. In practice it is possible to calculate the ϕ-values at these levels from those at the standard levels 200, 500, and 850 mb, by adding standard ϕ differences. Thus $\phi(833) \cong \phi(850) +$ standard $(\phi(833) - \varphi(850))$. If a different choice of levels is wanted, one has merely to introduce a monotonic transformation $\sigma = \sigma(p)$ in the derivation of the equations (3). The new standard levels are then found by solving

$$\sigma(p_k) = \frac{k - {}^1/_2}{n} [\sigma(p_0) - \sigma(0)]$$

for p_k. The transformation $\sigma = p^2$ is particularly convenient. Taking $p_0 = 1000$ mb again, one finds $p_1 = 408$ mb, $p_2 = 707$ mb, and $p_3 = 913$ mb, and one may use data at the standard levels 400, 700, and 900 mb. These give better definition to the low-level motions.

3. Method of Solving the Prediction Equation.—To solve the equations (3) and (4) we take as given the fields of ϕ_k and q_k at times $t - \Delta t$ and t at the points of the rectangular net $x = i\Delta s$, $y = j\Delta s$ $(i = 0, \ldots, l;\ j = 0, \ldots, m)$. The local time derivative $\partial q_k/\partial t$ at time t is evaluated from a finite-difference analog of (3) and q_k at time $t + \Delta t$ is obtained from

$$q_k(t + \Delta t) = q_k(t - \Delta t) + 2\Delta t \frac{\partial q_k}{\partial t}(t).$$

The values of $\phi_k(t + \Delta t)$ are then found by solving the finite-difference analog of the system (4),

$$(\Delta s)^{-2}(\phi_{i+1jk} + \phi_{i-1jk} + \phi_{ij+1k} + \phi_{ij-1k} - 4\phi_{ijk}) + \sum_{r=1}^{n} A_{rk}\phi_{ijr} = q_{ijk}(t + \Delta t)$$

in which the subscripts i and j denote quantities at $x = i\Delta s$, $y = j\Delta s$. The solution of this equation may conveniently be obtained by the iterative scheme

$$\phi_{ijk}^{(\nu+1)} = \phi_{ijk}^{(\nu)} + \lambda[4 - (\Delta s)^2 B_{kk}]^{-1}\left\{\phi_{i+1jk}^{(\nu)} + \phi_{i-1jk}^{(\nu+1)} + \phi_{ij+1k}^{(\nu)} + \phi_{ij-1k}^{(\nu+1)} - 4\phi_{ijk}^{(\nu)} + (\Delta s)^2\left[\sum_{r=1}^{n} A_{rk}\phi_{ijr}^{(\epsilon)} - q_{ijk}(t + \Delta t)\right]\right\}$$

$$(i = 1, \ldots, l-1;\ j = 1, \ldots, m-1;\ k = 1, 2, 3);$$
$$(\nu = 0, 1, 2, \ldots;\ \epsilon = \nu \text{ for } r \geq k,\ \epsilon = \nu + 1 \text{ for } r < k)$$

applied to the points ijk in lexicographic order with k first, j second, and i last.[3] It may be shown that the sequence $\{\phi_{ijk}^{(\nu)}\}$ will converge whatever the initial guess $\phi_{ijk}^{(0)}$ providing only that the matrix (A_{rk}) is negative definite and λ is appropriately chosen.

A 24-hour prediction carried out on the IAS electronic computer in 48 half-hour time steps Δt for a 16×16 grid required 48 minutes of computing time. During this time the machine performed approximately 750,000 multiplications and divisions, 10,000,000 additions and subtractions, and executed 30,000,000 distinct orders. The space interval Δs used for all predictions was 300 km.

4. Presentation of Results.—At 12:30 Z (Greenwich Mean Time), November 24, 1950, a small depression was observed to form on a cold front extending along the Atlantic seaboard of the United States and to develop within 24 hours into one of the worst storms ever recorded in the eastern United States. The surface weather sequence is shown in figures 1 (*a*, *b*, *c*, *d*). Because of the spontaneity and intensity of its development this storm was selected as an ideal test case for the prediction of cyclogenesis. The previous integrations with E_1 and E_2 had been made for the six overlapping 24-hour periods 11/23/03Z–11/24/03Z, 11/23/15Z–11/24/15Z, 11/24/03Z–11/25/03Z, 11/24/15Z–11/25/15Z, 11/25/03Z–11/26/03Z, and 11/25/15Z–11/26/15Z. The model E_1 gave predictions at 500 mb, and the model E_2 predictions at 300 and 700 mb. The predictions with E_3 were made for the same periods and at first for the levels 200, 500, and 850 mb.

It is not possible to say with precision whether or not cyclogenesis was predicted with the autobarotropic model; this model yields only a 500 mb forecast, whereas most of the development, in the sense of increase in cyclonic vorticity and kinetic energy, took place at lower levels. It may be seen from figure 5, however, that the changes that did occur at 500 mb during the 24 hours following 15Z, November 24, were poorly predicted.

The charts in figure 2 show that a deepening did actually occur at 700 mb but was not predicted by the two-level model, which instead predicted a filling. However, it should be remarked that a rudimentary deepening was predicted in another place, just off the coast of the United States between 35° and 40° latitude, and that this may perhaps indicate that the mechanism of cyclogenesis is inherent, although in a crude form, even in the two-level model.

The actual motion at 850 mb resembled the surface motion; a new low center formed to the southwest of the original center after 12 hours and in the succeeding

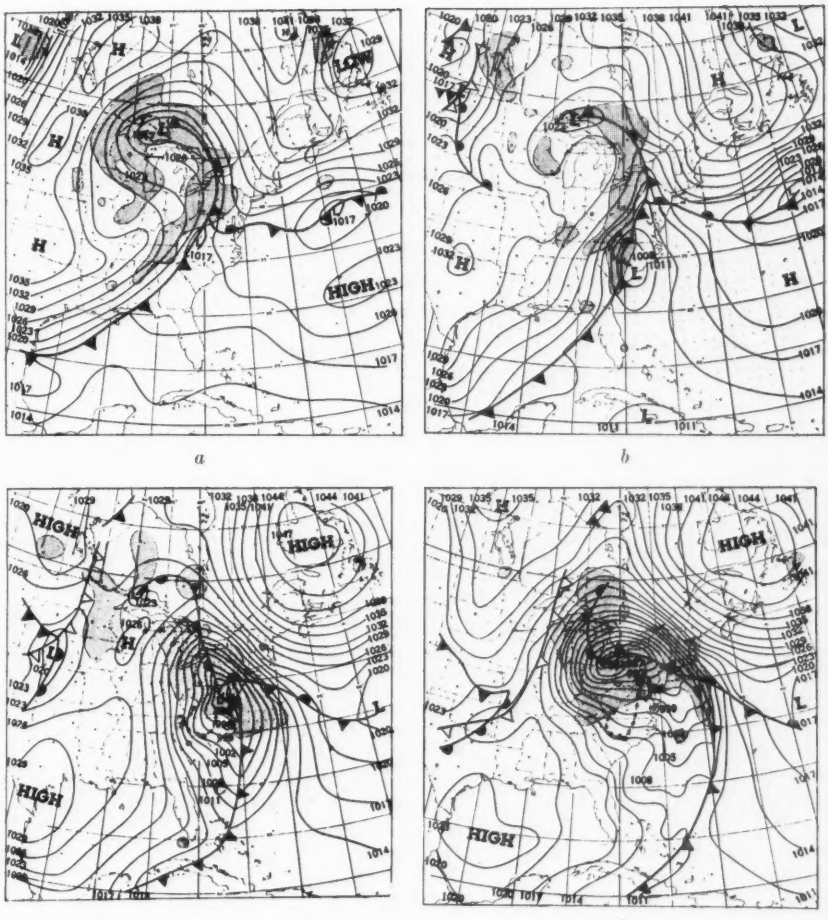


FIGURE 1

(*a*) Surface weather chart for 12:30 Z November 24, 1950. Shading indicates areas of active precipitation. (*b*) Surface weather chart for 00:30 Z November 25, 1950. Small squares connected by arrows indicate past positions of main low center at 12-hour intervals. (*c*) Surface weather chart for 12:30 Z November 25, 1950. (*d*) Surface weather chart for 00:30 Z November 26, 1950.

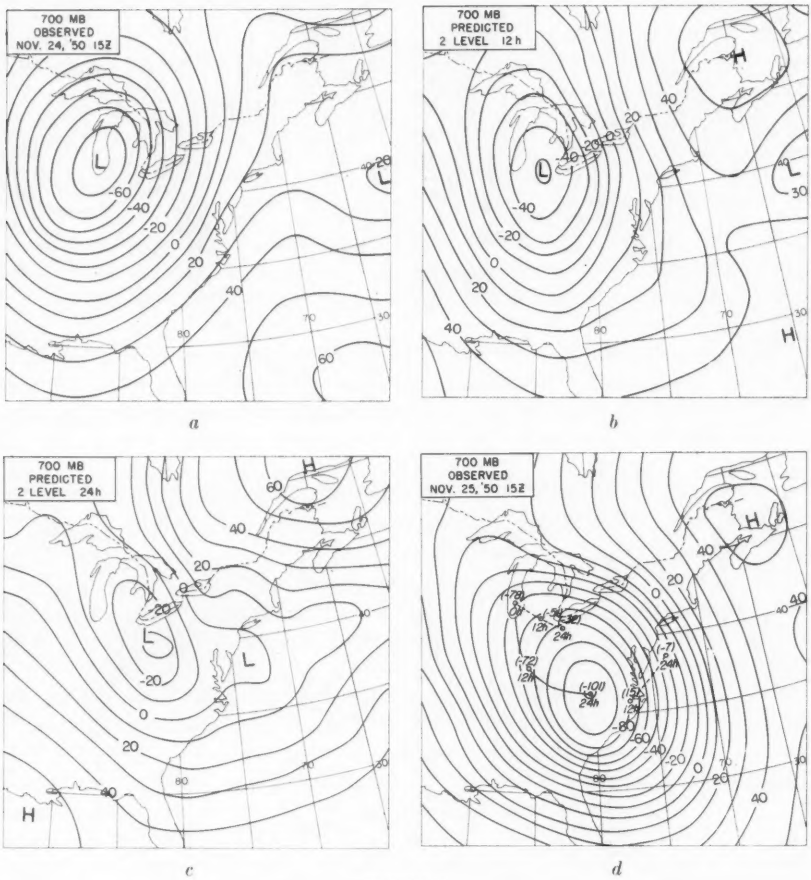


FIGURE 2

(*a*) Observed 700 mb height contours at 15 Z November 24, 1950. The contours are labeled as deviations from the standard height of 9879 ft. in units of 10 feet. (*b*) Two-level 700 mb prediction for 03 Z November 25. (*c*) Two-level 700 mb prediction for 15 Z November 25. (*d*) Observed 700 mb chart for 15 Z November 25. The small circles connected by solid lines indicate the successive positions of the observed low center, and those connected by dashed lines the predicted positions. The height difference at the center is printed above, and the time below, each circle.

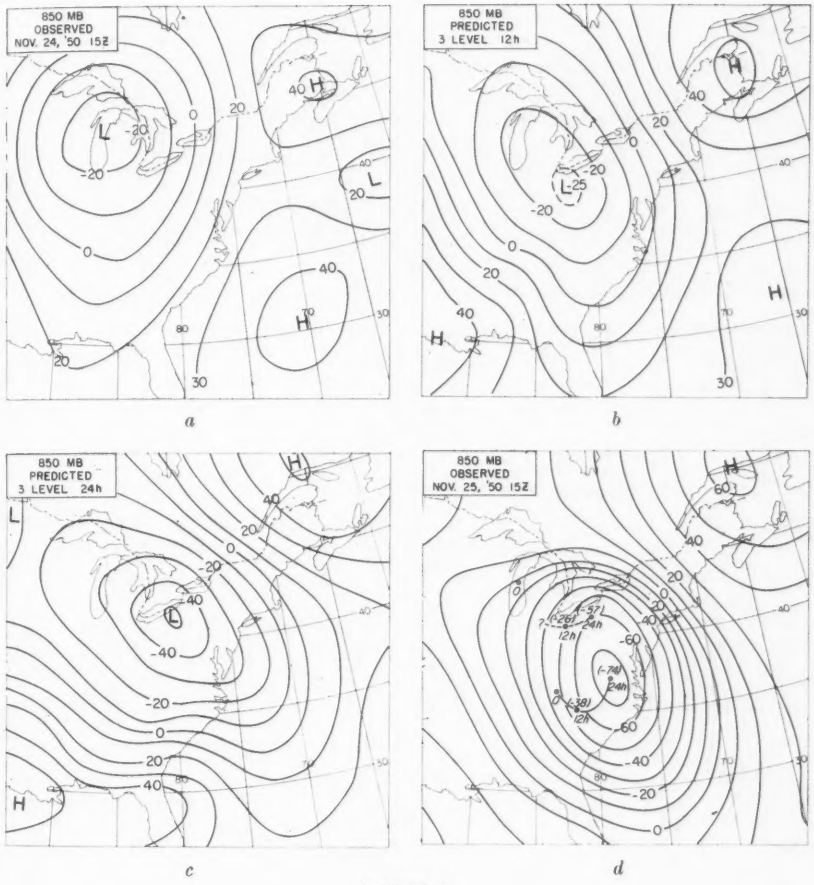


FIGURE 3

(*a*) Observed 850 mb height contours at 15 Z November 24, 1950. The contours are labeled as deviations from the standard height of 4779 ft. in units of 10 feet. (*b*) Two-level 850 mb prediction for 03 Z November 25. (*c*) Three-level 850 mb prediction for 15 Z November 25. (*d*) Observed 850 mb chart for 15 Z November 25. See legend under figure 2 for explanation of circles and connecting lines.

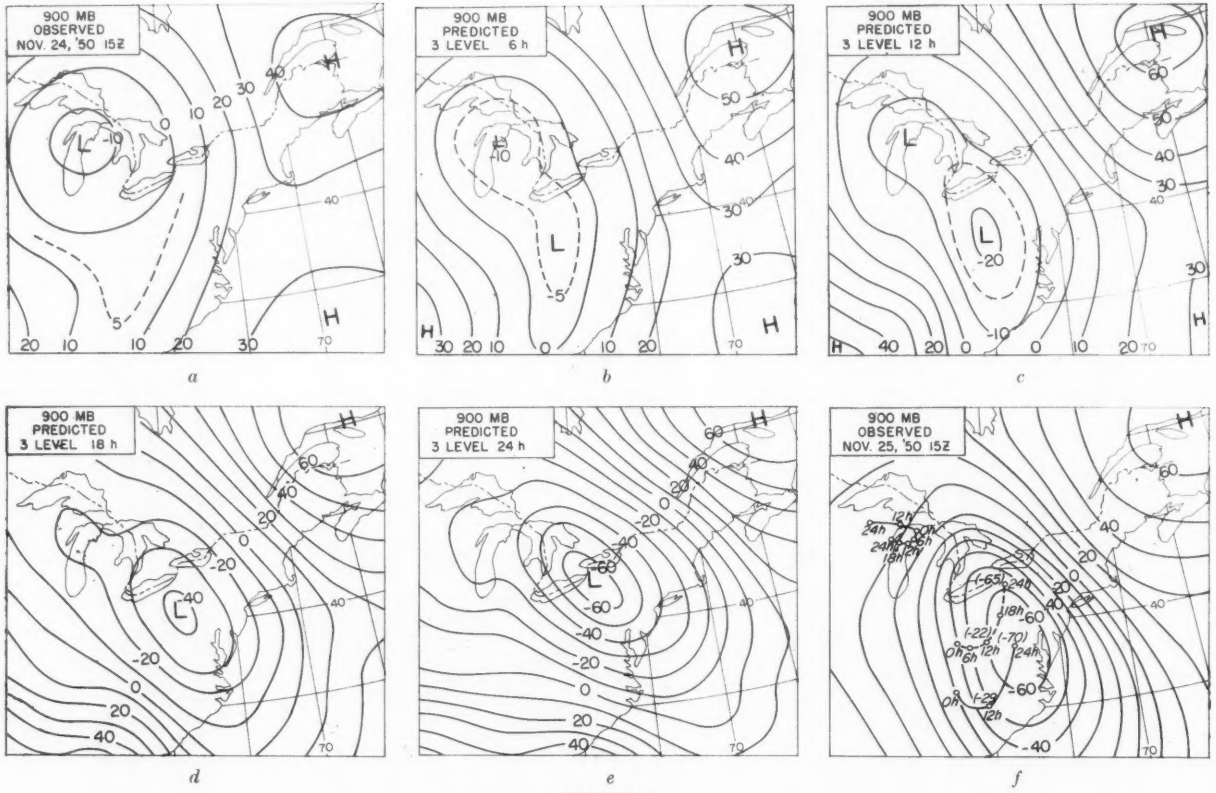


FIGURE 4

(*a*) Observed 900 mb height contours at 15 Z November 24, 1950. The contours are labeled as deviations from the standard height of 3242 ft. in units of 10 feet. (*b*) Three-level 900 mb prediction for 21 Z November 24. (*c*) Three-level 900 mb prediction for 03 Z November 25. (*d*) Three-level 900 mb prediction for 09 Z November 25. (*e*) Three-level 900 mb prediction for 15 Z November 25. (*f*) Observed 900 mb chart for 15 Z November 25. See legend under figure 2 for explanation of circles and connecting lines.

12 hours grew to such an extent that it dominated the entire flow over the eastern United States. It will be seen from figure 3 that a good part of this development was predicted by the three-level model, although as far as could be ascertained no separate low center was predicted to form.

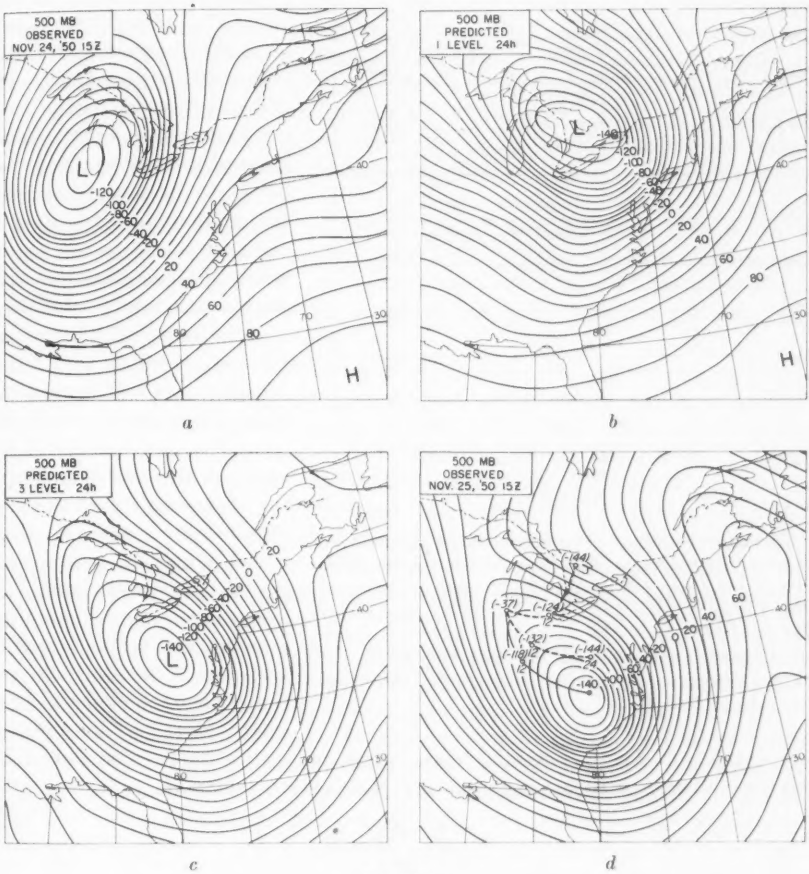


FIGURE 5

(*a*) Observed 500 mb height contours at 15 Z November 24, 1950. The contours are labeled as deviations from the standard height of 18281 ft. in units of 10 feet. (*b*) One-level (autobarotropic) 500 mb prediction for 15 Z November 25. (*c*) Three-level 500 mb prediction for 15 Z November 25. (*d*) Observed 500 mb chart for 15 Z November 25. See legend under figure 2 for explanation of circles and connecting lines. The upper dashed line is the path of the low center in the one-level model, the lower dashed line is the path in the three-level model.

The failure to predict a new center can be attributed in part to the location of the forecast levels in the three-level model. The 850 and 500 mb geopotentials do not adequately describe the motion between the ground and say 700 mb, yet it was here that the most intense part of the development took place. An additional

computation was therefore made with a three-level model defined for the 900, 700, and 400 mb levels. The results are shown in figure 4. It will be seen that a new center was predicted to form and to deepen by very nearly the correct amount. The path of the predicted center (dashed trajectory in figure 4*f*) paralleled the observed path (solid trajectory) but remained some 240 miles too far north.

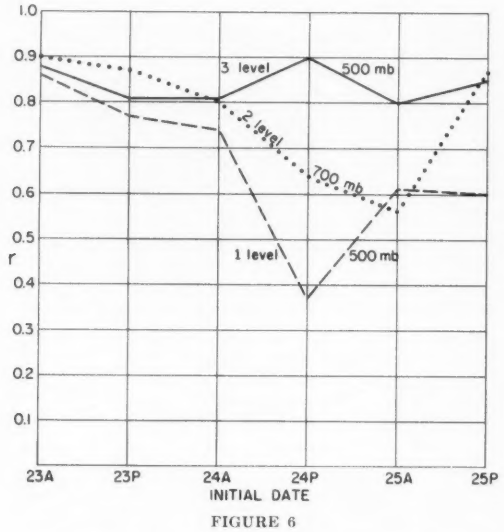


FIGURE 6

Correlation coefficient r between predicted and observed 24-hour height changes plotted as a function of the initial time, for the period November 23–25, 1950. The letters *A* and *P* denote the times 03 Z and 15 Z, respectively. The dashed line gives r for the one-level prediction at 500 mb, the dotted line gives r for the two-level prediction at 700 mb, and the continuous line gives r for the three-level prediction at 500 mb.

The discrepancy in the placement of the low center is believed to be associated with the spurious intensification of the anticyclone to the southwest. Since recalculation with shifted boundaries indicated that the intensification was not a boundary effect, and measurements showed impossibly large anticyclonic vorticities, it was decided that the intensification was probably due to the following mathematical simplification made in the model equations, a simplification that destroys the mechanism preventing the unlimited growth of anticyclonic vorticity: If η is the sum $f + \zeta$ of the vertical components of the earth's vorticity f and the relative vorticity ζ, θ is the potential temperature, and d/dt is the particle derivative, the basic quasi-geostrophic equation of motion may be written[2]

$$\frac{d}{dt}\left(\eta \frac{\partial \theta}{\partial p}\right) = \dot{0},$$

or

$$\frac{1}{\eta}\frac{d\eta}{dt} + \left(\frac{\partial \theta}{\partial p}\right)^{-1} \frac{d}{dt}\frac{\partial \theta}{\partial p} = 0.$$

It follows immediately that if η is once positive at a particle it cannot subsequently become negative, i.e., ζ cannot become smaller than $-f$. But if, as in the derivation of equations (3) and (4), the denominators η and $\partial\theta/\partial p$ are replaced by their respective mean values at the levels p_k, one no longer restricts the algebraic value of η, which may then become negative or impossibly small.

That the three-level 200-500-850 mb prediction considerably bettered the one-level, or autobarotropic, prediction can be seen from figure 5. This figure compares

the 500 mb 24-hour predicted charts from the initial time 1500Z November 24, with the observed chart for the same time. The position and orientation of the low as well as of the down-stream wedge are seen to be much more accurate in the three-level 500 mb forecast.

An over-all comparison of the one-, two-, and three-level 24-hour predictions was made by computing the correlation coefficient, r, between the predicted and observed 24-hour changes for each of the models. The coefficients were computed for the one-level predictions at 500 mb, the two-level predictions at 700 mb, and the three-level predictions at 500 mb. The graphs in figure 6 show the respective coefficients plotted against the initial times. It will be observed that the level of accuracy for all three predictions was high during the period preceding the storm, i.e., until 15Z November 24, but remained high only in the case of the three-level predictions.

5. Conclusions.—Following the predictions of the November 1950 storm a second series was carried out with the three-level model for a more typical, though less intense, development on the east coast of the United States. Again the cyclogenesis was predicted. With this result as additional evidence the following tentative conclusions may be drawn:

(1) The inception and development of cyclones are determinate and predictable events, consequently the conception of cyclogenesis as the growth of an effectively random perturbation in an unstable uniform field of flow must be abandoned. If the notion of hydrodynamic instability is to be retained at all, some such interpretation as the following would seem to fit the facts more closely: Since the basic flow is not steady and linear but transient, curved, and with variable curvature, the initial perturbation will have a preferred location in space and time, and its amplitude, though it may be small initially, will be entirely determined by the basic flow. It is like an automobile being pushed slowly but inexorably over a cliff.

(2) Since the models used were quasi-geostrophic, the geostrophic approximation is evidently not incompatible with the occurrence of cyclogenesis. Hence motions which are essentially non-geostrophic need not be assumed for the explanation of cyclogenesis. Thus, the Helmholtz-Solberg-Höiland type of instability that occurs under certain conditions for axially-symmetric displacements of the fluid in a cylindrical vortex, or constant transverse displacements in a quasi-linear flow—displacements in which angular and linear momentum respectively are conserved—decidedly does not appear responsible for cyclogenesis.

(3) Since the models contained no sources or sinks of heat we infer that these are not essential factors in the occurrence of extratropical cyclogenesis in at least a wide class of motions. This is not to say that they play no role at all, but that they do not play a decisive role.

(4) The evidence drawn from the two-level predictions indicates that two parameters along the vertical do not sufficiently define the motion for an accurate prediction of cyclogenesis. It appears that one must somehow take into account the low-level thermal asymmetries, and for this one needs at least a three-parameter model.

Acknowledgment.—The author wishes to thank Dr. Bruce Gilchrist who, assisted by Mr. Glen Lewis, programmed and carried out the three-level computations on

the Institute for Advanced Study Electronic Computer, and Mr. E. Hövmoller who analyzed the charts used in the forecasts.

* This work was sponsored jointly by O.N.R. and the Geophysics Researah Directorate of the Air Force Cambridge Research Center, under contract N6-ori-139 with O.N.R. It was presented at a joint meeting of the American Meteorological Society and the Fluid Dynamics Section of the American Physical Society in Washington, D. C., May 1953.

[1] Charney, J. G., and Phillips, N. A., "Numerical Integration of the Quasi-geostrophic Equations for Barotropic and Simple Baroclinic Flows," *J. Meteorol.*, **10**, No. 2, 71–99 (1953).

[2] Charney, J. G., "On the Scale of Atmospheric Motions," *Geofys. Publikasjoner, Oslo*, **17**, No. 2, 17 pp. (1948).

[3] This iteration method is essentially the extrapolated Liebmann method. See Frankel, S. P., "Convergence Rates of Iterative Treatments of Partial Differential Equations," *Math. Tables and Other Aids to Computation*, **4**, No. 30, 65–75 (1950); and Young, D., "Iterative Methods for Solving Partial Difference Equations of Elliptic Type" (1951). Seminar given at Numerical Analysis Seminar of Ballistic Research Laboratories, Aberdeen Proving Ground, Md. (based on Ph.D. thesis by author, Harvard University, Cambridge, Mass., 1950).

ARITHMETIC GENERA AND THE THEOREM OF RIEMANN-ROCH FOR ALGEBRAIC VARIETIES

By Friedrich Hirzebruch

THE INSTITUTE FOR ADVANCED STUDY

Communicated by S. Lefschetz, December 21, 1953

Introduction.—In a preceding note[1] we posed the problem whether the arithmetic genus $\Pi(V_n)$ of a (non-singular) algebraic variety[2] coincides with the Todd genus $T(V_n)$. The purpose of the present note is to prove that this is actually the case. Moreover we prove a main theorem (M) which gives a formula for the Euler-Poincaré characteristic of V_n with respect to the cohomology[3] of V_n with coefficients in the sheaf (faisceau) of local holomorphic cross-sections of any complex analytic bundle W over V_n which has the complex vector space C_q as fibre and the linear group $GL(q, C)$ as structure group. The main theorem expresses this Euler-Poincaré characteristic as a polynomial in the Chern classes of the tangential bundle of V_n and in the Chern classes of the bundle W. As special cases one gets the Todd formula $\Pi(V_n) = T(V_n)$ and the "Theorem of Riemann-Roch for arbitrary dimensions." The author wishes to extend his hearty thanks to Professors A. Borel, K. Kodaira, and D. C. Spencer, with all of whom he had many valuable discussions. The notes of Kodaira and Spencer[4] are essentially used. The author also wants to point out that the main theorem (in a slightly different formulation) was conjectured by J. P. Serre in a letter to Kodaira and Spencer. The proof of the main theorem uses the index theorem of the author[1] which involves essentially the theory of "cobordisme" due to R. Thom. Full details of the proof of the main theorem will appear elsewhere.

1. The Main Theorem and Some of Its Consequences.—Let V_n be an algebraic variety (non-singularly imbedded in some complex projective space). Let W be a complex analytic bundle over V_n with the complex vector space C_q as fibre and the linear group $GL(q, C)$ as structure group. Denote by c_i the Chern classes of the

tangential bundle of V_n and by d_j the Chern classes of the bundle W. ($c_i \in H^{2i}(V_n, Z)$, $d_j \in H^{2j}(V_n, Z)$, $0 \leqq i \leqq n$, $0 \leqq j \leqq q$, $c_0 = d_0 = 1$.) We introduce the formal roots γ_i, δ_j:

$$\sum_{i=0}^{n} c_i x^i = \prod_{i=1}^{n} (1 + \gamma_i x), \qquad \sum_{j=0}^{q} d_j x^j = \prod_{j=1}^{q} (1 + \delta_j x).$$

Every formal power series which is symmetric in the γ_i as well as in the δ_j will be considered as a power series in the Chern classes c_i and d_j and hence as an element of the cohomology ring of V_n. Denoting by Q the rationals we define the operator κ_n on the cohomology ring $H^*(V_n, Q)$ as follows: For every $u \in H^*(V_n, Q)$ we take the component of topological dimension $2n$ which we consider in the unique fashion as a rational number. This number is denoted by $\kappa_n[u]$.

MAIN THEOREM (M). *Using the above notations we put*

$$\chi(V_n, W) = \sum_{i=0}^{n} (-1)^i \dim H^i(V_n, W),$$

where $H^i(V_n, W)$ is the i-dimensional cohomology group of V_n with coefficients in the sheaf (faisceau) of local holomorphic cross-sections of W. We have

$$\chi(V_n, W) = \kappa_n \left[(e^{\delta_1} + e^{\delta_2} + \ldots + e^{\delta_q}) \prod_{i=1}^{n} \frac{-\gamma_i}{e^{-\gamma_i} - 1} \right] =$$

$$\kappa_n \left[e^{c_1/2} (e^{\delta_1} + e^{\delta_2} + \ldots + e^{\delta_q}) \prod_{i=1}^{n} \frac{\gamma_i/2}{\sinh \gamma_i/2} \right].$$

Remark: $x/\sinh x$ is a power series in x^2. Since the γ_i^2 can be regarded as the formal roots of the Pontrjagin polynomial[1] of V_n, we see that $\chi(V_n, W)$ is a polynomial in c_1, the Pontrjagin classes of V_n, and the Chern classes of the bundle W.

Now let W be the trivial bundle over V_n with C_1 as fibre. The main theorem implies[5]

THEOREM 1. *The Todd genus of an algebraic variety is equal to the arithmetic genus:*

$$\Pi(V_n) = \sum_{i=0}^{n} (-1)^i g_i = T(V_n).$$

More generally, let F be an arbitrary complex line bundle (fibre C_1) over V_n. Denote by $f = c(F)$ its characteristic class ($f \in H^2(V_n, Z)$). The main theorem implies

THEOREM 2. *We have*

$$\chi(V_n, F) = \kappa_n \left[e^{f + c_1/2} \prod_{i=1}^{n} \frac{\gamma_i/2}{\sinh \gamma_i/2} \right].$$

Hence $\chi(V_n, F)$ is a polynomial in $f + c_1/2$ and the Pontrjagin classes of V_n.

Theorem 2 can be considered as a generalization of the theorem of Riemann-Roch to arbitrary dimensions. It contains the known Riemann-Roch theorems for $n = 1, 2, 3$. In the case $n = 2$ the term $\dim H^1(V_n, F)$ has to be identified with the superabundance of F. A similar remark applies to $\dim H^1(V_n, F)$ and $\dim H^2(V_n, F)$ in the case $n = 3$. In general, the term $\dim H^n(V_n, F)$ is equal to

dim $H^0(V_n, K - F)$, where K is the canonical bundle of V_n. The characteristic class of K is equal to $-c_1$. Kodaira[6] proved that $H^i(V_n, F) = 0$ for $i > 0$, if the characteristic cohomology class of the bundle $F - K$ "contains" a closed positive definite Hermitian form. In this case $F - K$ is called ample. An arbitrary divisor D of V_n represents a complex line bundle,[7] which we also denote by D. Obviously dim $|D| + 1 =$ dim $H^0(V_n, D)$. Denoting the cohomology class of D by d we get

THEOREM 2*. *If $D - K$ is ample (in the sense of Kodaira), then we have*

$$\dim |D| + 1 = \kappa_n \left[e^{d + c_1/2} \prod_{i=1}^{n} \frac{\gamma_i/2}{\sinh \gamma_i/2} \right].$$

(For $n = 1$, this is the classical formula: dim $|D| = d - p$, where $c_1 = 2 - 2p$.)

Applying (M) to the case where W is the bundle of complex covariant p-vectors, yields[5]

THEOREM 3. *Let $h^{p,q}$ be the number of harmonic forms of type (p, q) on V_n. We have*

$$\chi^p(V_n) = \chi(V_n, W) = \sum_{q=0}^{n} (-1)^q h^{p,q} = \kappa_n \left[\left(\sum e^{-(\gamma_{i_1} + \ldots + \gamma_{i_p})} \right) \cdot \prod_{i=1}^{n} \frac{-\gamma_i}{e^{-\gamma_i} - 1} \right].$$

The sum on the right side has to be extended over all the $\binom{n}{p}$ possible combinations.

We attach to every algebraic variety the polynomial

$$\chi_y(V_n) = \sum_{p=0}^{n} \chi^p(V_n) y^p$$

and get by easy calculations

THEOREM 3*. *We have*

$$\chi_y(V_n) = \kappa_n \left[\prod_{i=1}^{n} Q_y(\gamma_i) \right],$$

where

$$Q_y(x) = \frac{-x(y+1)}{e^{-x(y+1)} - 1} - yx.$$

Theorem 3* is for $y = 0$ a restatement of Theorem 1. For $y = -1$ it gives that

$$\chi_{-1}(V_n) = \sum_{p,q=0}^{n} (-1)^{p+q} h^{p,q}$$

is equal to the Euler-Poincaré characteristic c_n of V_n. For $y = 1$ we obtain

THEOREM 4. *We have*

$$\chi_1(V_n) = \sum_{p,q=0}^{n} (-1)^q h^{p,q} = \kappa_n \left[\prod_{i=1}^{n} \frac{\gamma_i}{\operatorname{tgh} \gamma_i} \right].$$

Theorem 4 is known even for Kähler manifolds, because the left side of the equation is equal to the index $\tau(V_n)$.[8] The right side is also equal to $\tau(V_n)$ by the index theorem of the author.[1] The main theorem admits many further applications which we do not mention in this note.

2. A Sketch of the Proof of the Main Theorem.—In section 1 we derived the Theorems 1–4 from the main theorem (M). Actually the proof of (M) goes almost in the opposite direction and involves the following steps.

2.1. Proof of Theorem 1 by Means of the Known Theorem 4.—A complex manifold M_m is called a split manifold, if the structure group $GL(m, C)$ of its tangential bundle can be reduced complex analytically to the group $\Delta(m, C)$ of all triangular matrices (all entries above the diagonal are 0). Let V_n be an arbitrary (non-singular) algebraic variety, the tangential bundle has $GL(n, C)$ as structure group. Hence we can construct the associated bundle with $GL(n, C)/\Delta(n, C)$ as fibre. We obtain in this way an algebraic variety V_m^* of dimension $m = n + {}^1/_2 n(n-1)$. The fact that V_m^* is algebraic can be proved, for example, by using the general theorem of Kodaira[9] that every complex projective bundle over an algebraic variety is an algebraic variety. It can be proved that V_m^* is a split manifold. It is almost obvious that $\Pi(V_m^*) = \Pi(V_n)$. On the other hand, it follows from Theorem 4.2 in the preceding note[1] that $T(V_m^*) = T(V_n)$. Hence it suffices to prove Theorem 1 for all *algebraic* split varieties. Now let V_m^* be an arbitrary algebraic split variety. Then we have over V_m^* an increasing sequence of bundles W_i (in the sense of section 1), where W_i has C_i as fibre and where W_m is the tangential bundle of V_m^* ($0 \leqq i \leqq m$). We denote the complex line bundle W_i/W_{i-1} by A_i ($1 \leqq i \leqq m$) and its cohomology class by a_i. We use the notation of the preceding note and obtain readily the formula[10]

$$2^m T(V_m^*) = \sum_{j=0}^{m} \sum \tau(a_{r_1}, a_{r_2}, \ldots, a_{r_j}). \tag{1}$$

The inner sum has to be extended always over all the possible $\binom{m}{j}$ combinations. By the four-term formula of Kodaira-Spencer[11] we define in the faisceau theory virtual indices χ_1. This can be done for all algebraic varieties. (It can be done not only for χ_1, but even for χ_y.) We know that $\chi_1 = \tau$ for all non-singular algebraic varieties, (Theorem 4). Since χ_1 and τ both fulfill the "functional equation" 3.3 of the preceding note,[1] we get by an induction argument on the dimension that χ_1 and τ also coincide in the virtual case.[12]

By some rather complicated calculations based on exact sequences of sheaves we can prove the following formula which corresponds to (1)

$$2^m \Pi(V_m^*) = \sum_{j=0}^{m} \sum \chi_1(A_{r_1} \circ A_{r_2} \circ \ldots \circ A_{r_j}) \tag{2}$$

Here the terms χ_1 denote virtual indices.[13] Formulae (1) and (2) imply Theorem 1 for split varieties, and hence Theorem 1 is proved in general.

2.2. Proof of Theorem 2.—The virtual Π and the virtual T both fulfill the "functional equation" 4.3 of the preceding note. Hence the usual induction argument already used for χ_1 in 2.1 proves that Π and T also coincide in the virtual case. This proves Theorem 2 (i.e., the main theorem for $q = 1$). Moreover, the main theorem follows immediately for all cases in which the bundle W admits the triangular group $\Delta(q, C)$ as structure group.

2.3. Proof of the Main Theorem.—We use the notation of the first section. We construct over V_n the bundle V_m^* with fibre $GL(q, C)/\Delta(q, C)$ which is associated

with W. The manifold V_m^* is, according to Kodaira's theorem,[9] algebraic ($m = n + {}^1/_2q(q-1)$). We lift the bundle W up to V_m^* and call the lifted bundle W^*. The bundle W^* splits, i.e., admits $\Delta(q, C)$ as structure group. Hence Theorem 2 implies that $\chi(V_m^*, W^*)$ is given correctly by the formula of the main theorem. A spectral sequence argument shows that $\chi(V_m^*, W^*) = \chi(V_n, W)$. On the other hand, we can prove by a slight generalization of the Theorem 4.2 of the preceding note[1] that the formula of the main theorem gives for $\chi(V_m^*, W^*)$ and $\chi(V_n, W)$ the same values. This concludes the proof of the main theorem.

Remark.—The formulae (1) and (2) are special cases of more general formulae valid for χ_y with y regarded as indeterminate. The "functional equations" for $\Pi = \chi_0 = T$ and for $\chi_1 = \tau$ are special cases of a "functional equation" valid for χ_y.

[1] These PROCEEDINGS, **39**, 951–956 (1953). We use the notations of the preceding note.

[2] Complex dimensions are indicated by a subscript.

[3] All the cohomology groups occurring in this note have finite dimensions. See Cartan, H., and Serre, J. P., *Compt. rend. acad. sci., Paris*, **237**, 128–130 (1953), and Kodaira, K., these PROCEEDINGS, **39**, 865–868 (1953).

[4] Kodaira, K., and Spencer, D. C., these PROCEEDINGS, **39**, 641–649, 868–877 (1953). K. Kodaira, *loc. cit.* in ref. 3.

[5] Here we use the theorem of Dolbeault (*Compt. rend. acad. sci., Paris*, **236**, 175–177 (1953)) which states that $H^q(V_n, W^{(p)}) \cong H^{p,q}$, where $W^{(p)}$ is the bundle of covariant p-vectors and where $H^{p,q}$ is the linear space of all harmonic forms of type (p, q) on V_n.

[6] Kodaira, K., "On a Differential-Geometric Method in the Theory of Analytic Stacks," these PROCEEDINGS, **39**, 1268–1273 (1953).

[7] Conversely every complex line bundle can be represented by a divisor. See Kodaira, K., and Spencer, D. C., *loc. cit.* in ref. 4, p. 874.

[8] Hodge, W. V. D., *Proc. Int. Congr. Math.*, **I**, 182–192 (1952), and *Proc. Lond. Math. Soc.* (*3*), **1**, 104–117 (1951).

[9] Kodaira, K., not yet published.

[10] The sum on the right side starts for $j = 0$ with $\tau(V_m^*)$. The formula (1) was used by the author to prove Theorem 4.1 of the preceding note (*loc. cit.* in ref. 1). See also Hirzebruch, F., "The Index of an Oriented Manifold and the Todd Genus of an Almost Complex Manifold," Notes, Princeton University, 1953 (mimeographed).

[11] Kodaira, K., and Spencer, D. C., "On a Theorem of Lefschetz and the Lemma of Enriques-Severi-Zariski," these PROCEEDINGS, **39**, 1273–1278 (1953), formula (14).

[12] The fact that χ_1 fulfills the functional equation can be obtained from the four-term formula by some calculations.

[13] The small circle $\circ$ denotes "virtual intersection."

ON THE CONTINUOUS GOLD-MINING EQUATION

BY RICHARD BELLMAN AND SHERMAN LEHMAN

THE RAND CORPORATION AND STANFORD UNIVERSITY

Communicated by J. von Neumann, December 30, 1953

§1. *Introduction.*—In some previous communications[1, 2] we have described some results obtained in the investigation of a dynamic programming problem, the "gold-mining" problem, which led to the functional equation

$$f(x, y) = \text{Max}\begin{bmatrix} \text{A: } \sum_{i=1}^{N} p_i(r_i x + f((1 - r_i)x, y) \\ \text{B: } \sum_{i=1}^{N} q_i(s_i y + f(x, (1 - s_i)y) \end{bmatrix}, \qquad x, y \geq 0, \qquad (1.1)$$

where $p_i, q_i \geq 0$, $\sum_i p_i, \sum_i q_i < 1$. The solution of this equation was given, *inter alia*, in ref. 1, and shown to have a relatively simple form. In addition, a partial solution of a more complicated equation, corresponding to a non-linear utility function, was given in ref. 2, having the same form. It is known, however, as a result of an unpublished counter-example due to H. N. Shapiro and S. Karlin, that the solution of more general equations such as

$$f(x, y) = \text{Max}\begin{bmatrix} \text{A: } p_1(r_1 x + f((1 - r_1)x, y) \\ \text{B: } q_1(r_2 y + f(x, (1 - r_2)y) \\ \text{C: } p_2(r_2 x + r_4 y + f((1 - r_3)x, (1 - r_4)y) \end{bmatrix}, \qquad (1.2)$$

$0 \leq r_1, r_2, r_3, r_4 \leq 1$, $0 < p_1, q_1, p_2 < 1$ cannot have the same simple form for all values of the parameters.

In an effort to gain some insight into the structure of particular classes of solutions of (1.2), and similar equations of more complicated type, we have been led to consider some continuous analogues of these equations. There are many different procedures for obtaining these continuous analogues. One which we have followed to begin with leads to problems in the calculus of variations.

An essential feature of our research lies in viewing a policy in its extensive rather than normal form, to borrow the terminology of game theory. Another way of stating this is that instead of determining the complete solution for one set of initial parameters, which would correspond to determining the extremal curve in the classical theory of the calculus of variations, we attack our problem by imbedding it into the family of problems of this type with arbitrary initial parameters. This is the approach used throughout the theory of dynamic programming.[1-5] Having done this we determine an optimal continuation from each position, which upon being carried through yields an optimal policy.

This approach, which may be considered a variant in problems of deterministic type, is in many ways a necessity in problems of stochastic type. It is possible to treat many of the classical problems in the calculus of variations by means of this technique. We shall return to this point at some future time.

Guided by our knowledge of the solution in the discrete cases, and using the be-

havioristic approach described above, we have been able to solve completely and explicitly a variety of problems which are intractable in the original discrete form.

In the following sections we shall discuss the simplest counterpart of (1.2) in continuous form, and list a number of typical results we have obtained. Following this we shall sketch briefly a formulation of the more general continuous version which results from processes corresponding to (1.1) and which requires more powerful techniques.

A more complete discussion and proofs of the results contained herein will appear elsewhere. Further results concerning more general problems discussed in ref. 3 will be presented subsequently.

§2. *Formulation.*—In the formulation of problems involving the use of continuous policies we are immediately faced with the difficulty of defining what we mean by a continuous mixed strategy, and of constructing the appropriate mathematical theory with which to handle this thorny concept. To circumvent these conceptual and mathematical difficulties, we shall utilize an idea emphasized in ref. 6, which—briefly put—is that for mathematical purposes, mixing at a point is to an arbitrary degree of approximation equivalent to mixing pure strategies in an interval about the point.

Let us then consider a process where we are given two initial quantities, x and y, the gold mines of ref. 2; and two operations, A and B, mining operations. If A is used over a time interval δ, there is a probability $1 - q_1\delta + o(\delta)$ that $r_1x\delta + o(\delta)$ is obtained and that the process is allowed to continue, with the new initial amounts $x - r_1x\delta + o(\delta)$, y; and a probability $q_1\delta + o(\delta)$ that nothing is obtained and the process terminates. In a like manner, if B is used, there is a probability $1 - q_2\delta + o(\delta)$ that $r_2y\delta + o(\delta)$ is obtained and the process continues; and a probability $q_2\delta + o(\delta)$ that the process terminates.

To introduce the concept of mixing, we consider first the case in which the time interval is divided into intervals of length Δ, where Δ is small. In a typical interval $[t, t + \Delta]$, $t = k\Delta$, the first part of the interval, $[t, t + \phi_1\Delta]$, will be devoted to the use of A; while the second part, $[t + \phi_1\Delta, t + \Delta]$, will be devoted to the use of B. In the limit, as $\Delta \to 0$, we obtain the effect of mixing A and B at t in the ratio $\phi_1:(1 - \phi_1)$, cf. ref. 6 for further discussion.

A strategy consists of a choice of ϕ_1 for each of the points $k\Delta$. We wish to determine the strategy which will maximize the expected value of the amount obtained before the process terminates. For any given strategy let

$$\begin{aligned} x(t) &= \text{quantity of gold remaining in first mine provided that the process has continued until } t, \\ y(t) &= \text{quantity of gold remaining in second mine provided that the process has continued until } t, \\ p(t) &= \text{probability that the process continues at least to } t, \\ f(t) &= \text{expected amount obtained up to } t. \end{aligned} \tag{2.1}$$

Writing down the equations expressing $x(t + \Delta)$, $y(t + \Delta)$, $p(t + \Delta)$, $f(t + \Delta)$ in terms of the values at t, and letting $\Delta \to 0$, we are led to the following system of differential equations:

$$\begin{aligned}
\frac{dx}{dt} &= -\phi_1(t)r_1x(t), & x(o) &= x_0,\\
\frac{dy}{dt} &= -\phi_2(t)r_2y(t), & y(o) &= y_0,\\
\frac{dp}{dt} &= -p(t)[\phi_1(t)q_1 + \phi_2(t)q_2], & p(o) &= 1,\\
\frac{df}{dt} &= p(t)[\phi_1(t)r_1x(t) + \phi_2(t)r_2y(t)], & f(o) &= 0,
\end{aligned} \tag{2.2}$$

where $0 \leq \phi_1 \leq 1$, $\phi_2 = 1 - \phi_1$. The problem is now to determine $\phi_1(t)$ so as to maximize $f(\infty)$. It is not difficult to give a proof based upon, say, weak convergence, which will assure us that the maximum is actually attained. As W. Fleming has kindly informed us, the existence of a maximum is guaranteed by a general theorem in the calculus of variations.

Since the equations are fortunately non-linear, variational techniques are particularly applicable. We find

THEOREM 1. *The maximum value of $f(\infty)$ is attained by the policy*

(a) *If* $q_2r_1x > q_1r_2y$, $\phi_1 = 1$,
(b) *If* $q_1r_2y > q_2r_1x$, $\phi_2 = 1$,
(c) *If* $q_2r_1x = q_1r_2y$, $\phi_1 = r_2/(r_1 + r_2)$, $\phi_2 = r_1/(r_1 + r_2)$. (2.3)

Note that the boundary line is, as might be expected, the set of points where expected gain over expected cost is the same for both choices, A and B.

THEOREM 2. *If T is finite, the optimal policy has one of the following six forms:*

(a) *A always* (d) *A, then M, then A*
(b) *B always* (e) *B, then M, then A*
(c) *M followed by A* (f) *B followed by A.* (2.4)

This is for $q_1 < q_2$; a similar result holds for $q_2 < q_1$. The precise intervals within which each is used may be determined explicitly. Here M represents the choice given in (2.3c).

The optimal strategy represents a compromise between the long-term policy given in Theorem 1 and the short-term policy of maximizing expected gain.

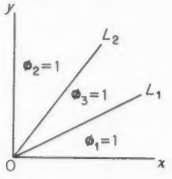
The 3-choice problem corresponding to (1.2) has the continuous analogue:

$$\begin{aligned}
\frac{dx}{dt} &= -[\phi_1(t)r_1 + \phi_3(t)r_3]x(t),\\
\frac{dy}{dt} &= -[\phi_2(t)r_2 + \phi_3(t)r_4]y(t),\\
\frac{dp}{dt} &= -p(t)[\phi_1(t)q_1 + \phi_2(t)q_2 + \phi_3(t)q_3],\\
\frac{df}{dt} &= p(t)[(\phi_1(t)r_1 + \phi_3(t)r_3)x(t) + (\phi_2(t)r_2 + \phi_3(t)r_4)y(t)], \quad f(o) = 0,
\end{aligned} \tag{2.5}$$

where $0 \leq \phi_1, \phi_2, \phi_3 \leq 1$, $\phi_1 + \phi_2 + \phi_3 = 1$.

The maximum value of $f(\infty)$ is provided in the general case by the policy repre-

sented schematically by



(2.6)

Depending upon the values of the parameters, one line is an absorbing barrier, which is to say, for (x, y) on the line a mixed policy is pursued which keeps the point on the line. This line has as its equation the equality of expected gain over expected cost. The other line will be a translucent barrier, causing only a change from $\phi_i = 1$ to $\phi_j = 1$, and is not defined by an equality of the above type. In special cases the middle region disappears and L_1 coincides with L_2. The solution is now that for the two-choice case.

This last result is quite surprising and explains some of the difficulties of the discrete problem. One boundary, the absorbing barrier, is determined by a local condition, whereas the other is determined by a global condition.

THEOREM 3. *If, in the two-choice problem described by* (2.2), *in place of expected return, we seek to maximize the expected value of some function* ϕ *of the total return where* ϕ *is any strictly increasing function, the solution is that given by* (2.3).

To obtain this result we consider

$$G = -\int^{\infty} \phi(x_0 + y_0 - x(t) - y(t))dp(t), \tag{2.7}$$

the quantities being defined as in (2).

The proofs of the above results are long and detailed, depending upon a precise analysis of the properties of an optimal policy.

§3. *More General Processes.*—Let us now consider the more general process corresponding to (1.1). Here the use of A leads to a variety of possible gains, and similarly for B. The quantities $x(t)$ and $y(t)$, as defined by (2.1) are now stochastic quantities. This means that it is no longer possible to obtain the equations of (2.2). Instead we must introduce the function $F(u, v, t)$ defined by the property that

$$\Pr\{u \leq x(t) \leq u + du, v \leq y(t) \leq v + dv\} = F(u, v, t)dudv. \tag{3.1}$$

We may now, in a way similar to that followed in §2, derive a partial differential equation for F of the form

$$\frac{\partial F}{\partial t} = P(u, v)\frac{\partial F}{\partial u} + Q(u, v)\frac{\partial F}{\partial v}. \tag{3.2}$$

The system of ordinary differential equations

$$\frac{du}{dt} = P(u, v), \qquad \frac{dv}{dt} = Q(u, v) \tag{3.3}$$

connected with (3.2) will have a form similar to the first two equations in (2.2).

The differential equations we have used to define our continuous processes bear the same relation to the rigorous integral equations defined by the original processes as the heat equation bears to the Chapman-Kolmogoroff equations.

Finally, let us note that the above formalism is also applicable to two-person multi-stage games of continuous type, and, in particular, to pursuit games.

These extensions will be discussed in subsequent communications.

[1] Bellman, R., "On the Theory of Dynamic Programming," PROC. NATL. ACAD. SCI., **38**, 716–719 (1952).

[2] Bellman, R., "Some Functional Equations in the Theory of Dynamic Programming," *Ibid.*, **39**, 1077–1082 (1953).

[3] Bellman, R., "Bottleneck Problems and Dynamic Programming," *Ibid.*, **39**, 947–951 (1953).

[4] Bellman, R., "A Problem in the Theory of Dynamic Programming," *Econometrica* (to appear).

[5] Bellman, R., "Computational Problems in the Theory of Dynamic Programming," *Proc. Symposium on Numerical Analysis*, Santa Monica, 1953.

[6] Bellman, R., and Blackwell, D., "Some Two-Person Games Involving Bluffing," PROC. NATL. ACAD. SCI., **35**, 600–605 (1949).

ON A PARTICULAR TYPE OF CONVERGENCE TO A SINGULAR MATRIX

BY E. INONU AND E. P. WIGNER

PRINCETON UNIVERSITY

Communicated December 14, 1953

It was attempted, in an earlier investigation,[1] to consider limiting cases of continuous groups G which are themselves groups. For this purpose, linear combinations of the infinitesimal elements I of G were formed

$$J_i = \sum(u_{i\nu} + \epsilon w_{i\nu})I_\nu \tag{1}$$

in which u was singular but the determinant $|u + \epsilon w|$ did not vanish identically in ϵ. It was attempted then to find the conditions for the J to remain infinitesimal operators of a group even in the limiting case $\epsilon \to 0$. The group generated by the J in the limiting case was called a contraction of G.

It was stated, in the course of the derivation of the conditions for the contraction, that it is always possible to find two non-singular matrices β and α such that

$$\beta u \alpha^{-1} = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix} \qquad \beta w \alpha^{-1} = \begin{Vmatrix} v & 0 \\ 0 & 1 \end{Vmatrix}. \tag{2}$$

In (2), the 1, 0, v are submatrices: the rank of the 1 in $\beta u \alpha^{-1}$ is equal to that of u. This statement is, in general, incorrect[2] as can be seen already on the example of the two dimensional matrices

$$u = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix} \qquad w = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}.$$

When checking the calculation which led to the form (2), it was found that it contained a continuity assumption which appears reasonable but which was not stated explicitly. Since the continuity assumption in question may be of interest also apart from the problem of the contraction of groups, we shall formulate it below.

Let us consider two n dimensional vector spaces I and J and mappings $U(\epsilon)$ of I on J. For $0 < \epsilon < \epsilon_0$, the mapping shall be one-to-one; $U(0)$ shall map I on a subspace J_u of J. We shall call the mapping $U(\epsilon)$ *reversibly continuous* at $\epsilon = 0$ if $U(0)U(\epsilon)^{-1}$, applied to the vectors of J_u, converges to the identical mapping of J_u on itself as $\epsilon \to 0$. An equivalent formulation hereto is that the mapping $U(0)$ $U(\epsilon)^{-1}U(0)$ of I unto J shall converge, as $\epsilon \to 0$, to $U(0)$. In fact, let i be a vector in I and write $j = U(0)i$. The convergence of $U(0)$ $U(\epsilon)^{-1}j \to j$ is synonymous with the convergence of $U(0)$ $U(\epsilon)^{-1}U(0)i$ to $U(0)i$.

If the convergence of $u + \epsilon w$ to u is reversibly continuous, the existence of the normal forms (2) does indeed follow. *Thus the addition of this assumption will guarantee the validity of the aforementioned considerations.*[1] One can see this perhaps most simply as follows. Let us denote by I_0 the subspace of I for which $uI_0 = 0$. Similarly, let us denote the subspaces uI and wI_0 by J_u and J_{0w}. We can first prove then that J_u and J_{0w} together span the whole space J. In fact, if the vector i_0 is in I_0, $wi_0 = 0$ would entail $(u + \epsilon w)i_0 = 0$ for all ϵ which is contrary to assumption. Furthermore, wi_0 cannot be contained in J_u because if it were

$$u(u + \epsilon w)^{-1}wi_0 = (1/\epsilon)u(u + \epsilon w)^{-1}(u + \epsilon w)i_0 = (1/\epsilon)ui_0 = 0$$

would not converge to wi_0, contrary to the postulate of reversibly continuous convergence. Hence, wI_0 contains as many linearly independent vectors as I_0 and these are also linearly independent of the vectors in uI. It follows that $uI = J_u$ together with wI_0 span the whole space J. If we choose, therefore, a coordinate system in I so that some of its axes i_{0k} span I_0, the others i_{1k} the rest of I, we shall have

$$ui_{0k} = 0 \qquad wi_{0k} = j_{0k} \qquad ui_{1k} = j_{uk}. \tag{3}$$

The last two of these are the definitions of j_{0k} and j_{uk} which, together, span J. We can set, therefore

$$wi_{1k} = \sum_\lambda v_{\lambda k}j_{u\lambda} + \sum_\lambda z_{\lambda k}j_{0\lambda}. \tag{3a}$$

Hence if we write

$$i_{uk} = i_{1k} - \sum z_{\lambda k}i_{0\lambda} \tag{3b}$$

the i_{uk} and i_{0k} will span I and we have, in addition to (3) which remain valid

$$ui_{uk} = j_{uk} \qquad wi_{uk} = \sum_\lambda v_{\lambda k}j_{u\lambda}. \tag{3c}$$

Hence u and w will in fact have the form (2) if we choose in I the coordinate axes i_{uk}, i_{0k} and in J the axes j_{uk}, j_{0k}.

In matrix language, one can formulate our theorem as follows. Be u and w n-dimensional square matrices and the rank r of u be smaller than n while $u + \epsilon w$ be non-singular except for isolated ϵ. Then the necessary and sufficient condition for

$$u(u + \epsilon w)^{-1}u \to u \qquad \text{as } \epsilon \to 0 \tag{4}$$

is that the coefficient of ϵ^{n-r} in the determinant $|u + \epsilon w|$ shall not vanish. In order to prove this, we first note that (4) remains invariant if one substitutes $\beta u\alpha^{-1}$ for u and $\beta w\alpha^{-1}$ for w where α and β are non-singular and independent of ϵ. Ac-

cording to our Theorem, if (4) is fulfilled, β and α can be so chosen that (2) be fulfilled. This form shows, however, that the lowest term in the determinant $|\beta u\alpha^{-1} + \beta w\alpha^{-1}|$, if considered as a polynomial in ϵ, is ϵ^{n-r}. Hence, the lowest term in the determinant $|u + \epsilon w|$ is $|\beta|^{-1}|\alpha|\epsilon^{n-r}$, i.e., the coefficient of ϵ^{n-r} in the determinant $|u + \epsilon w|$ does not vanish.

Conversely, let us assume that u and w are of dimension n, u of rank r, and the coefficient of ϵ^{n-r} in $|u + \epsilon w|$ does not vanish. The same holds then for $u' = \beta u\alpha^{-1}$ and $w' = \beta w\alpha^{-1}$ and we can choose β and α so that $u' = \beta u\alpha^{-1}$ have the form given in (2). Furthermore, the algebraic complement of any element in the first r rows of $|u' + \epsilon w'|$ goes to zero at least as ϵ^{n-r}. Hence, the ratio of such a complement to the determinant $|u' + \epsilon w'|$, i.e., any element in the first r rows of $(u' + \epsilon w')^{-1}$, converges to a finite number (or to 0) as $\epsilon \to 0$. It follows that the elements in the first r rows of $(u' + \epsilon w')^{-1}u'$ converge to the same numbers to which the corresponding elements of $(u' + \epsilon w')^{-1}(u' + \epsilon w')$ converge. Hence, as far as the first r rows are concerned

$$u'(u' + \epsilon w')^{-1}u' \to u'(u' + \epsilon w')^{-1}(u' + \epsilon w') = u'. \tag{4a}$$

However (4a) is clearly correct also as far as the last $n - r$ rows are concerned because these vanish on both sides because of the form (2) of $u' = \beta u\alpha^{-1}$. Hence, (4a) follows from the premises and since α and β are independent of ϵ, this holds also for (4).

[1] Inonu, E., and Wigner, E. P., these PROCEEDINGS, **39**, 510 (1953).

[2] We are much indebted to Dr. V. Bargmann for his friendly comments concerning this point.

---

*ON NON-NEGATIVE-VALUED MATRICES**

BY Y. K. WONG

PRINCETON UNIVERSITY

Communicated by J. von Neumann, December 11, 1953

Consider a matrix $A = (a_{ij})$ of m rows and n columns with non-negative elements. Define

$$s_j^{(k)} = \sum_{i=1}^{k} a_{ij} \qquad (j, k = 1, 2, \ldots, n). \tag{1}$$

We let $s_j = s_j^{(n)}$ for short, and let $\|A\| = \max(s_1, \ldots, s_n)$. In particular, for a row vector α' with non-negative coordinates $a_1, \ldots, a_n$, $\|\alpha'\| = \max_i(a_i)$. If β is a column vector with non-negative coordinates $b_1, \ldots, b_m$, we have $\|\beta\| = \sum_i b_i$. The definition for $\|A\|$ has the properties of a "norm" in Banach space. We use $A(., k)$ to denote the kth column of A.

Two square matrices are said to be equivalent to each other if one can be transformed into the other by permutations of the same rows and columns.

THEOREM. *Let $A = (a_{ij})$ be a square matrix of order n with non-negative elements such that $s_j \leqq 1$ for $j = 1, 2, \ldots, n$. Then the following properties are equivalent to one another. (i) $I - A$ is non-singular. (ii) The matrix A is equivalent to one satis-*

fying $s_k^{(k)} < 1$ *for* $k = 1, 2, \ldots, n$. (*iii*) *For some* $h < n$ *and* $e > 0$, *the matrix* A *satisfies* $\|\sigma' A^h\| \leqq e < 1$, *where* $\sigma' = (s_1, \ldots, s_n)$. (*iv*) $\lim_q \|A^q\|^{1/q} < 1$. (*v*) *The matrix* A *is equivalent to one satisfying* $\sum_{i=1}^{k-1} s_i^{(k)} a_{ik} < 1 - a_{kk}$ *for* $k = 2, 3, \ldots, n$, *and* $0 < 1 - a_{11}$.

We shall show that the conditions (*i*), (*ii*), (*iii*), (*iv*) imply one another in cyclic order, and (*ii*) implies (*v*), (*v*) implies (*i*). (*i*) *implies* (*ii*): If all the sums s_k were equal to 1, then $1 - \sum_i a_{ik} = 0$ for $k = 1, \ldots, n$, which would be the same as $(1, 1, \ldots, 1)(I - A) = 0$. In other words, the n rows of $I - A$ would be linearly dependent, and $I - A$ would be singular. Thus, condition (*i*) implies that one sum, say s_n, is less than 1. Now $d_n = \det(I - A) > 0$ implies that $d_k > 0$, where d_k is the determinant of the submatrix $I - A_k$ consisting of the first k rows and columns of $I - A$ (ref. 3, p. 237, Corollary 2). Thus $d_k > 0$ implies the existence of a column in A_k such that the sum of its elements is less than 1. Let such a column be the kth one of A_k, where $k = n - 1, n - 2, \ldots, 1$. This proves that (*i*) implies (*ii*). (*ii*) *implies* (*iii*): If every element below the principal diagonal, a_{ik} with $i > k$, are zeros, then by (*ii*), we have $s_j = \sum_1^n {}_i a_{ij} = \sum_1^j {}_i a_{ij} < 1$ for $j = 1, 2, \ldots, n$. This implies that $\|A\| < 1$ and hence the conclusion. Consider now a_{ik} for $i > k$ are not all zeros. If on the $(n - 1)$th column, $a_{n,\,n-1} = 0$, then $s_{n-1}^{(n)} = s_{n-1}^{(n-1)} < 1$ by (*ii*). We, thus, have two columns $A(\cdot, n - 1)$, $A(\cdot, n)$ with sums $s_{n-1} < 1$, $s_n < 1$. If $a_{n,\,n-1} > 0$, then by (*ii*), $s^{(n)} < 1$ and $s^{(n)} a_{n,\,n-1} < a_{n,\,n-1}$. It follows that $\sigma' A(\cdot, n - 1) = \sum s_i^{(n)} a_{i,\,n-1} < \sum_i a_{i,\,n-1} \leqq 1$ by hypothesis. In either case, we have $\sigma' A(\cdot, n - 1) < 1$. If either $a_{n,\,n-2}$ or $a_{n-1,\,n-2}$ is not zero, then $\sigma' A(\cdot, n - 2) < 1$ or $\sigma' A^2(\cdot, n - 2) < 1$, respectively. Hence, $\sigma' A^2(\cdot, j) < 1$ for $j = n - 2, n - 1, n$. Continuing this type of reasoning, we can determine an integer $h \leqq n - 1$ such that $\sigma' A^h(\cdot, j) < 1$ for all $j = 1, 2, \ldots, n$. Consequently, there exists an e such that $0 < e < 1$ and $\|\sigma' A^h\| \leqq e$. This proves that (*ii*) implies (*iii*). (*iii*) *implies* (*iv*): Let $a_p = \log \|\sigma' A^{p-1}\| = \log \|A^p\|$. Then $a_0 = 0$ and $a_{p+q} \leqq a_p + a_q$. Pólya and Szegö (ref. 2, Problem 98) show that $\lim_p (a_p/p) = \inf_p (a_p/p)$. By hypothesis, we have $a_p < 0$ if $p - 1 \geqq h$. Hence $a_p/p < 0$ for all $p \geqq h + 1$, and $b = \inf_p (a_p/p) < 0$. It follows that $\lim_p \|A^p\|^{1/p} = e^b < 1$. In fact, e^b is 0 or $0 < e^b < 1$ according to $b = -\infty$ or $b > -\infty$. This proves that (*iii*) implies (*iv*). That (*iv*) implies (*i*) is well known in Banach algebras. Evidently, (*v*) follows from (*ii*). Finally, that (*v*) implies (*i*) follows from inequality (9) in (ref. 4, p. 139). This completes the proof of the theorem.

Another obvious equivalent condition is that all the proper values of A lie in a unit circle with center at the origin.

A matrix A is said to be quasi-nilpotent if $\lim \|A^q\|^{1/q} = 0$. For a finite square matrix, quasi-nilpotency implies nilpotency and conversely.

COROLLARY. *Let* A *be a square non-negative matrix of order* n *such that* $s_j \leqq 1$ *for* $j = 1, 2, \ldots, n$. *If* A *is not nilpotent, and has a simple proper value,* λ_1, *with the greatest modulus, then*

$$0 < \lambda_1 = \lim_q |A^q\|^{1/q} = e^b, \qquad b = \inf_p [p^{-1} \cdot \log \|A^p\|] > -\infty.$$

For, by Frobenius' result,[1] λ_1 is non-negative. Since A is not nilpotent, $\lambda_1 > 0$ As λ_1 is simple by hypothesis, it follows, from Syvester's spectral representation of

a matrix, that, for sufficiently large p,

$$\lambda_1 - \epsilon \leqq \frac{\|A^{p+1}(.,k)\|}{\|A^p(.,k)\|} = \lambda_1 + \epsilon,$$

where $k = 1, 2, \ldots, n$. We can show that $(\lambda_1 - \epsilon)\|A^p\| \leqq \|A^{p+1}\| \leqq (\lambda_1 + \epsilon)\|A^p\|$. This means that $\lambda_1 = \lim \|A^{p+1}\|/\|A^p\| = \lim \|A^p\|^{1/p}$, which proves our result.

Frobenius shows that an indecomposable, non-negative matrix has a simple, positive maximum proper value. Thus, the Theorem of Vivanti and Pringsheim on power series with non-negative coefficients may be extended to a power series $\sum_r A^r t^r$ with the powers of an indecomposable non-negative matrix A.

Let A be a square matrix of order n with non-negative elements. Then t (real or complex) is in the resolvent set of A if $tI - A$ is regular. Let $B = |t|^{-1}A$; we may apply the preceding Theorem to $I - B$. The convergence of $\sum|t|^{-k-1}A^k$ implies that of $\sum t^{-k-1}A^k$. For simplicity, we consider $t > 0$. If A and B are square matrices with non-negative elements and satisfy the second Hilbert resolvent equation with respect to t, then every pair of matrices of the form $\alpha A + (1 - \alpha)B$, where $0 \leqq \alpha \leqq 1$, satisfies also the second resolvent equation with respect to t. The proof makes use of property (*ii*) in the above Theorem by verifying $t^{-1}\sum_{i=1}^{k}(\alpha \cdot a_{ik} + (1 - \alpha)a_{ik}) < 1$ for $0 \leqq \alpha \leqq 1$. This proves the convexity property of the set of non-negative matrices satisfying the second resolvent equation.

Applications.—We assume the regularity of $I - A$. If $\|A\| < 1$, then

$$\frac{s_j}{1 - \|A\|_l} \leqq \sum_{i=1}^{n} A_*(i,j) \leqq \frac{s_j}{1 - \|A\|}, \tag{2}$$

where $A_* = A + A^2 + \ldots + A^p + \ldots$, and $\|A\|_l = \min(s_1, \ldots, s_n)$.

The preceding inequalities are not useful in applications, as $\|A\| = 1$ arises frequently in practical problems. To overcome the difficulty arising from the unit modulus, we apply property (*iii*) in the above Theorem to determine an integer $h < n$ such that $\|\sigma' A^h\| < 1$, where $\sigma' = (s_1, \ldots, s_n) = (1, 1, \ldots, 1)A = \zeta' A$. We can verify that

$$\frac{\sigma'(I + A + \ldots + A^h)}{1 - \|\sigma' A^h\|_l} \leqq \zeta' A_* \leqq \frac{\sigma'(I + A + \ldots + A^h)}{1 - \|\sigma' A^h\|}. \tag{3}$$

Inequalities (3) are very significant for computation, as h is 1 or 2 in practical problems. When $h = 0$, the above two inequalities reduces to (2). If $\|A\| < 1$, then (3) holds for every h, and hence they may serve as refinements of (2).

Let $\beta' = (b_1, \ldots, b_n)$ be a vector with non-negative coordinates. Assuming the existence of $(I - A)^{-1}$, we have $(I - A)^{-1} = I + A_*$. With the aid of (2) or (3), we estimate the elements of $\beta' A_*$ and hence of $\beta'(I - A)^{-1} = \beta'(I + A_*)$. The approximations for $\beta' A_*$ are applicable to some expressions involving the resolvent of A by noting that $t(tI - A)^{-1} = I + (t^{-1}A)_*$ for $t > 0$. Moreover, inequality relations concerning the elements on each row of the resolvent $(tI - A)^{-1}$, $t > 0$, can be determined. Since $A_* - A = A_* A = A A_*$, we have

$$\beta' A_* = \beta'(I + A_*)A = \beta' A + \beta' A A_* \leqq \|\beta'\|\sigma' + \|\beta' A\|\zeta' A_*, \tag{4}$$

$$\beta' A_* = \beta' A(I + A_*) \leqq \|\beta' A\| \zeta'(I + A_*) = \|\beta' A\|(\zeta' + \zeta' A_*), \quad (5)$$

$$\beta' A_* \geqq \|\beta' A\|_\iota(\zeta' + \zeta' A_*), \quad (6)$$

$$\beta' A_* \geqq \|\beta' A\|_\iota \|A\|^{-1}\sigma' + \|\beta' A\|_\iota \zeta' A_*, \quad (7)$$

where $\|\beta' A\|_\iota = \min(\beta' A(.,1), \ldots, \beta' A(.,n))$. We obtain (7) from (6), as $\|A\|^{-1}\sigma' \leqq \zeta'$. Thus, (4) and (7) furnish an upper and a lower bound of similar nature. Since $\|A\|_\iota$ is usually very small, the value $\|\beta' A\|/\|A\|_\iota$ would be too large to render $\|\beta' A\| \cdot \|A\|_\iota^{-1} \cdot \sigma' + \|\beta' A\| \zeta' A_*$ a useful approximation for an upper bound. If we replace the moduli in (4) by the lower moduli, we would not secure a good lower bound in comparison with (6), since $\|\beta'\|_\iota = 0$ in practical problems. We may substitute (3) into (4)–(7). In case $\|A\| < 1$, we obtain by virtue of (2)

$$\|\beta' A\|_\iota \left(\zeta' + \frac{\sigma'}{1 - \|A\|_\iota}\right) \leqq \beta' A_* \leqq \|\beta' A\| \left(\zeta' + \frac{\sigma'}{1 - \|A\|}\right), \quad (8)$$

$$\|\beta' A\|_\iota \left(\frac{1}{\|A\|} + \frac{1}{1 - \|A\|_\iota}\right) \sigma' \leqq \beta' A_* \leqq \left(\|\beta'\| + \frac{\|\beta' A\|}{1 - \|A\|}\right) \sigma'. \quad (9)$$

The inequalities in (8) follow from (5) and (6), while those in (9) follow from (4) and (7). The inequalities (8) and the upper bound in (9) may be simplified as follows:

$$\frac{\|\beta' A\|_\iota}{1 - \|A\|_\iota} \zeta' \leqq \beta' A_* \leqq \frac{\|\beta' A\|}{1 - \|A\|} \zeta', \qquad \beta' A_* \leqq \frac{\|\beta'\|}{1 - \|A\|} \sigma'. \quad (10)$$

Observe that the lower and upper bounds for $\zeta' A_*$ given by (2) usually render poor estimates unless $\|A\|$ is sufficiently small. If

$$\|A\| < 1 - \|\beta'\|/c, \qquad c = \|\beta' + \sigma'\|, \quad (11)$$

then $\|\beta'\|/(1 - \|A\|) < c$, and the last inequality in (10) gives a better estimate than $\beta' A_* \leqq c\sigma'$, which is the same as inequality (30) in reference 3, page 239. In practical problems, c usually is at most 1. If (11) holds, then $c < 1$. For, if $c = 1$, it would imply that $c \leqq \|A\| + \|\beta'\| < 1$, contradicting to $c = 1$.

The inequalities (4)–(7) together with (3) are simple for practical problems. These inequalities will improve the lower and upper bounds for determinants in reference 3.

* The paper was prepared under the Navy contract N6onr-27009.

[1] Frobenius, G., "Über Matrizen aus nicht-negativen Elementen," *Sitzber. kgl. preuss. Akad. Wiss. zu Berlin*, pp. 456–477 (1912).

[2] Pólya and Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Vol. I, J. Springer, Berlin (1925).

[3] Wong, Y. K., "Some Inequalities of Determinants of Minkowski Type," *Duke Math. J.*, **19**, 231–241 (1952).

[4] Wong, Y. K., "An Inequality for Minkowski Matrices," *Proc. Am. Math. Soc.*, **4**, 137–141 (1953).

SOME RELATIONS FOR CRYSTALS WITH SUBSTRUCTURES

BY M. J. BUERGER

CRYSTALLOGRAPHIC LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated December 18, 1953

Introduction.—In many a complex crystal structure there can be recognized a set of atoms arranged in a pattern having simpler geometrical characteristics than the entire pattern. Such a set may be defined as a *substructure*. Familiar examples of this relation are found in crystals whose structures are based upon close packing of the larger atoms. Thus in kyanite, Al_2SiO_5, the oxygen atoms can be described as being arranged in cubic close packing. These atoms, considered alone, constitute a substructure which, ideally, has a smaller cell and higher symmetry than the complete structure of kyanite. The larger cell and reduced symmetry of the structure as a whole is due to the way the metal atoms are distributed among the interstices of the oxygen atoms. From a symmetry point of view, a crystal structure and its substructure (or substructures) conform to derivative structure theory:[1] the substructure is a basic structure while the entire crystal structure is a derivative structure, and has a derivative symmetry.

The presence of a substructure gives rise to specialized x-ray diffraction effects. Under favorable circumstances these effects are obvious and the existence of the substructure is accordingly revealed. In these cases, both the Patterson map and reciprocal lattice itself have specialized features which may be recognized.

Properties of the Vector Set of a Structure Containing a Substructure.—Let a set of points be decomposed into a subset and its complement subset. Then the vector set[2] of the entire set contains three vector subsets: the vector set of the subset, the vector set of the complement subset, and the cross products of the subset and complement subset. Since the substructure corresponds to a subset, it follows that the Patterson of a crystal with a substructure contains the Patterson of the substructure.

Now it is an elementary consequence of derivative-structure theory that the density of coincidence operations in the basic structure is a multiple of the density of coincidence operations in the derivative structure. Thus the weighting of the substructure peaks is a multiple of the weighting of the other peaks of the vector set, and therefore, other things being equal, the substructure peaks tend to be the most prominent features of the Patterson. Furthermore, in the commonest case, the primitive cell of the entire structure is a multiple, n, of the primitive cell of the substructure. The cells of the vector sets of these structures, of course, bear the same relation to one another.[2] This means that while one origin peak of the substructure coincides with the Patterson origin of the entire structure, the $(n - 1)$ intense origin peaks of the Patterson of the substructure occur within the cell of the Patterson of the entire structure. Thus, for a crystal containing a substructure, the most prominent feature of the Patterson is usually a set of strong peaks outlining the lattice of the substructure.

Conversely, the existence of a set of approximately equal peaks outlining a sublattice in a Patterson is evidence suggesting the existence of a substructure having that sublattice. For example, if one observes within the Patterson cell a set of

prominent peaks corresponding to a cubic face-centered lattice of appropriate translation, he has reason to suspect a substructure of atoms (presumably large electronegative atoms such as oxygen or sulfur) in cubic close packing.

One can go further than this. Since the Patterson of the substructure has the periodicity of the subcell, the vectors pertaining to the substructure are repeated along with the origins of the subcell. In other words, when the crystal contains a substructure its Patterson has a recognizable subperiodicity within the whole cell. It is possible to study this sub-Patterson intuitively, or it can be treated more systematically. For example, one can construct a minimum function[3, 4] for the n points of the sublattice, and let this function range the subcell. This minimum function lays down a value which approaches (i.e., is greater than or equal to) the Patterson of the substructure alone plus the cross terms, both of which have the periodicity of the subcell. This result can then be subtracted from the rest of the Patterson, leaving, as a residue, the Patterson of the complement structure alone. Since the substructure is usually easy to interpret, or even to guess, this may permit a complicated Patterson to be interpreted.

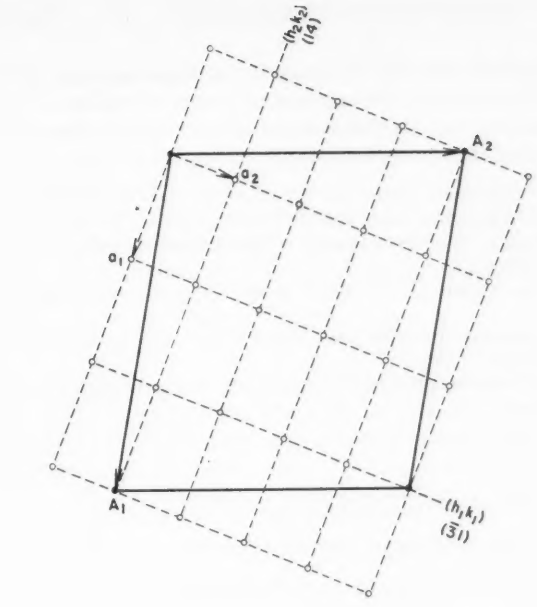
The Reciprocal Lattice of a Structure Containing a Substructure.—While the Patterson is probably the most powerful tool for investigating a crystal with a substructure, a good deal of information can be obtained from a mere inspection of a photograph of the reciprocal lattice[5] since, of course, the Patterson is derived from the arrangement and absolute weightings of the points of the reciprocal lattice. As pointed out above, the vector multiplicity in the substructure is comparatively high, and therefore the reciprocal terms contributing to the origins of the vector set of the substructure tend to be strong. This means that there should be heavily weighted points in the reciprocal lattice, especially near the origin, corresponding to the origin points of the substructure of the Patterson. Hence it is evident that although there is a small cell for the reciprocal lattice as a whole, yet certain points, especially near the origin, are prominent and mark out a super reciprocal cell. This supercell corresponds to the subcell (i.e., the cell of the substructure) in crystal space.

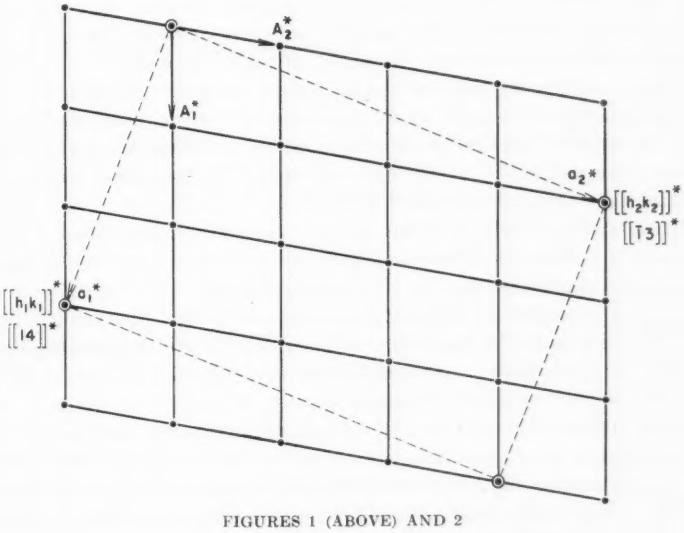
Figures 1 and 2 outline the relations involved in the lattices, sublattices, and superlattices of direct and reciprocal space. Figure 1 shows the lattice and sublattice in direct space, while figure 2 shows the corresponding cell and supercell in the reciprocal space. For simplicity the illustrations are two-dimensional.

In figure 1, the large cell A_1A_2 is the cell of the crystal. Within it is a substructure whose cell a_1a_2 is a submultiple ($^1/_{12}$ the area in the illustration) of the cell of the whole crystal. The subcells are delimited by planes (h_1k_1) and (h_2k_2). In the illustration these are taken as $(3\bar{1})$ and (14), respectively.

In figure 2 the cells reciprocal to those of figure 1 are shown. The cell reciprocal to A_1A_2 is the small cell $A_1{}^*A_2{}^*$. To find the cell reciprocal to the subcell a_1a_2 it is merely necessary to find the translations reciprocal to $d_{h_1k_1}$ and $d_{h_2k_2}$, specifically $[h_1k_1]^*$ and $[h_2k_2]^*$. Thus the cell in reciprocal space defined by the reciprocal lattice points $[[h_1k_1]]^*$ and $[[h_2k_2]]^*$ constitutes a supercell corresponding to the substructure, and these points may be expected to be prominent in reciprocal space.

Figures 1 and 2 are mutually reciprocal and can be interchanged; for example, figure 1 may be regarded as the reciprocal lattice picture from which the crystal lattice and its substructure of figure 2 can be derived. Note that one picture corre-





FIGURES 1 (ABOVE) AND 2

sponds to the projection of the crystal cell, the other to the photograph of the reciprocal lattice. The large cell of one is similar to[6] the small cell of the other if rotated 90° and vice versa.

Conclusions.—When a crystal contains a substructure the fact can often be de-

tected not only from its Patterson but also often from the photographs of its reciprocal lattice. The Patterson can be decomposed by means of a sublattice minimum function and manipulated so that the Patterson of the structure not concerned with the substructure is independently available for investigation.

[1] Buerger, M. J., "Derivative Crystal Structures," *J. Chem. Phys.*, **15**, 1–16 (1947).

[2] Buerger, M. J., "Vector Sets," *Acta Cryst.*, **3**, 87–97 (1950).

[3] Buerger, M. J., "Limitation of Electron Density by the Patterson Function," PROC. NATL. ACAD. SCI., **36**, 738–742 (1950).

[4] Buerger, M. J., "A New Approach to Crystal-Structure Analysis," *Acta Cryst.*, **4**, 531–544 (1951).

[5] Buerger, M. J., "Photography of the Reciprocal Lattice," *Am. Soc. for X-Ray and Electron Diffraction, Monograph Number 1* (1944).

[6] Buerger, M. J., *X-Ray Crystallography*, John Wiley & Sons, Inc., New York, 1942, p. 116.

---

*ON DIFFRACTION BY A STRIP**

BY A. ERDÉLYI AND C. H. PAPAS

CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated by P. S. Epstein, December 14, 1953

Introduction.—The problem of diffraction by an infinite strip or slit has been the subject of several investigations.[1] There are at least two "exact" methods for attacking this problem. One of these is the integral equation method,[2] the other the Fourier-Lamé method.[3] The integral equation obtained for this problem cannot be solved in closed form: expansion of the solution in powers of the ratio (strip width/wavelength) leads to useful formulas for low frequencies. In the Fourier-Lamé method the wave equation is separated in coordinates of the elliptic cylinder, the solution appears as an infinite series of Mathieu functions, and the usefulness of the result is limited by the convergence of these infinite series, and by the available tabulation of Mathieu functions.

The variational technique developed by Levine and Schwinger avoids some of the difficulties of the above-mentioned methods and, at least in principle, is capable of furnishing good approximations for all frequency-ranges. The scattered field may be represented as the effect of the current induced in the strip, and it has been proved by Levine and Schwinger[4] that it is possible to represent the amplitude of the far-zone scattered field in terms of the induced current in a form which is stationary with respect to small variations of the current about the true current. Substitution, in this representation, of a rough approximation for the current may give a remarkably good approximation of the far-zone scattered field amplitude. In this note we assume a normally incident field polarized parallel to the generators of the strip. As a rough approximation, we take a uniform density of the current induced in the strip. Since the incident magnetic field is constant over the strip, Fock's theory[5] may be cited in support of the uniformity of the current distribution, except near the edges where the behaviour of the field[6] indicates an infinite current density. A more detailed analysis of the current, by Moullin and Phillips,[3] is available but was not used here.

Once the (approximate) amplitude of the far-zone field has been obtained, the scattering cross-section may be found by the application of the scattering theorem[4, 7, 8] which relates this cross-section to the imaginary part of the amplitude of the far-zone scattered field along the central line of the umbral region. In spite of the crude approximation adopted for the induced current, the scattering cross-section shows a fair agreement with other available results.

Integral Equation.—We assume a plane wave with harmonic time dependence $\exp(-i\omega t)$, normally incident on the perfectly conducting infinitesimally thin strip of width $2a$,

$$z = 0, \qquad -a \leq y \leq a. \tag{1}$$

We further assume that the incident wave is polarized parallel to the edges of the strip (i.e., to the x axis) so that the only nonvanishing components of its electric field $\mathbf{E}$, magnetic field $\mathbf{H}$, and complex Poynting vector $\mathbf{S}$ (all measured in MKS units, bars denoting conjugate complex numbers) are

$$E_x^{\,i} = \exp(ikz) \tag{2}$$

$$H_y^{\,i} = k/(\omega\mu)\exp(ikz) \tag{3}$$

$$S_z^{\,i} = E_x^{\,i}\overline{H_y^{\,i}} = k/(\omega\mu) \tag{4}$$

The scattered electric field is again parallel to the x axis, and the total electric field is the sum of the incident and scattered field,

$$E_x(y, z) = E_x^{\,i}(y, z) + E_x^{\,sc}(y, z) \tag{5}$$

The boundary condition on the screen is the vanishing of the electric field

$$E_x(y, 0) = 0 \qquad \text{for } -a < y < a; \tag{6}$$

and the scattered field, $E_x^{\,sc}$, must represent, at large distances from the screen, an outgoing cylindrical wave (Sommerfeld's radiation condition).

The expression of the total electric field in terms of the induced current $K_x(y)$ is

$$E_x(y, z) = \exp(ikz) - \omega\mu/4 \int_{-a}^{a} H_0^{(1)}(k[(y - y')^2 + z^2]^{1/2})K_x(y')dy' \tag{7}$$

$i/4H_0^{(1)}(k[(y - y')^2 + z^2]^{1/2})$ being the two-dimensional free-space Green's function. The scattered field in (7) certainly satisfies Sommerfeld's radiation condition; in order that it also satisfies (6), we must have

$$1 = \omega\mu/4 \int_{-a}^{a} H_0^{(1)}(k|y - y'|)K_x(y')dy' \qquad |y| \leq a \tag{8}$$

and this is the integral equation of our problem.

Far-Zone Scattered Field Amplitude.—We define the far-zone scattered field amplitude $A(\phi)$ by

$$E_x^{\,sc} \sim i/4[2/(\pi ik\rho)]^{1/2}\exp(ik\rho)A(\phi) \qquad k\rho \to \infty \tag{9}$$

where $\rho\cos\phi = z$ and $\rho\sin\phi = -y$; this expression represents a cylindrical outgoing wave of "amplitude" A.

Since

$$H_0^{(1)}(k[(y - y')^2 + z^2]^{1/2}) \sim [2/(\pi ik\rho)]^{1/2} \exp(ik\rho + iky' \sin\phi) \quad (10)$$

when $k\rho \to \infty$, we have from (7), for large $k\rho$,

$$E_x^{sc} \sim i/4[2/(\pi ik\rho)]^{1/2} \exp(ik\rho) \int_{-a}^{a} i\omega\mu \exp(iky' \sin\phi) K_x(y')dy' \quad (11)$$

Comparing (9) and (11) we get

$$A(\phi) = i\omega\mu \int_{-a}^{a} \exp(iky' \sin\phi) K_x(y')dy' \quad (12)$$

A Scattering Theorem.—If P denotes the scattered energy flux per unit length, then the scattering cross-section σ is defined by

$$\sigma = P/S \quad (13)$$

where $S = 1/2 \text{ Re } S_z^i = k/(2\omega\mu)$ is the incident energy flux per unit area.

$$P = 1/2 \text{ Re} \int_{-a}^{a} E_x^i(y', 0)\bar{K}_x(y')dy' - 1/2 \text{ Re} \int_{-a}^{a} K_x(y')dy' \quad (14)$$

since $E_x^i = 1$ at $z = 0$. Hence,

$$\sigma = \omega\mu/k \text{ Re} \int_{-a}^{a} K_x(y')dy' \quad (15)$$

But from (12)

$$A(0) = i\omega\mu \int_{-a}^{a} K_x(y')dy' \quad (16)$$

Therefore, comparing (15) and (16) we get

$$\sigma = \text{Im } A(0)/k \quad (17)$$

It may be remarked that, in the limit for very high frequencies, the cross-section (17) turns out to be twice the projected area. At first sight the factor two may seem inconsistent with geometrical optics, but actually the discrepancy is due to the different definitions of the scattering cross section.[9]

Variational Principle.—We multiply the integral equation (8) by $K_x(y)$, integrate from $y = -a$ to $y = +a$, divide by the square of $\int_{-a}^{a} K_x(y)dy$, and then recall (16). Thus we obtain

$$A(0) = 4i \frac{\left[\int_{-a}^{a} K_x(y)dy\right]^2}{\int_{-a}^{a}\int_{-a}^{a} K_x(y)H_0^{(1)}(k|y - y'|)K_x(y')dy\,dy'} \quad (18)$$

It is easy to show that the expression (18) is stationary with respect to small variations of K_x about the true K_x which satisfies the integral equation (8). Consequently, if we substitute a reasonable trial function for the unknown K_x in (18), we expect to obtain a reasonably accurate value of $A(0)$.

Trial Function.—According to Fock's theory[5] the current induced in the central region of the strip is equal to twice the tangential component of the incident magnetic field. And according to Bouwkamp,[6] H_y must have a singularity like $(a^2 - y^2)^{-1/2}$ at the edges. This certainly conforms to the detailed information Moullin and Phillips[3] reported. In the present paper we take for a trial function $K_x(y) = 1$, thus ignoring the effect of the edges.

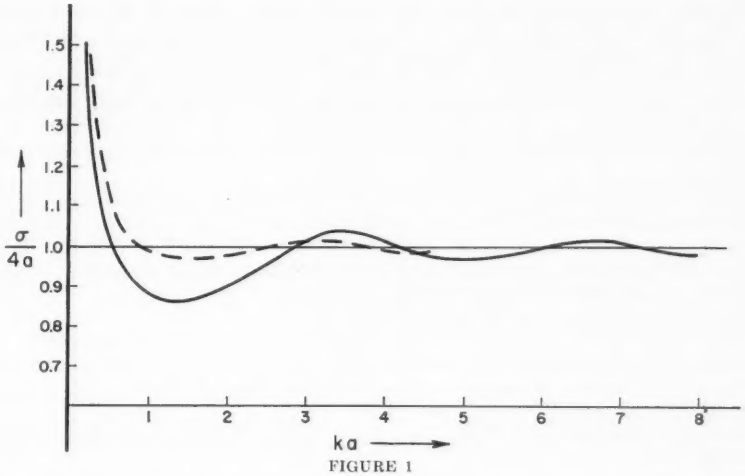


FIGURE 1

Plot of $\sigma/(4a)$ versus ka. Solid curve is variational solution. Broken curve is Morse and Rubinstein's.

Scattering Cross Section.—With $K_x(y) = 1$, computation shows that (18) becomes

$$[A(0)]^{-1} = (4iak)^{-1}\left\{\int_0^{2ka} H_0^{(1)}(t)dt - H_1^{(1)}(2ka) + (\pi ika)^{-1}\right\} \quad (19)$$

And applying the theorem (17) to (19), we get

$$\sigma/(4a) = \frac{\int_0^{2ka} J_0(t)dt - J_1(2ka)}{\left[\int_0^{2ka} J_0(t)dt - J_1(2ka)\right]^2 + \left[\int_0^{2ka} Y_0(t)dt - Y_1(2ka) - (\pi ka)^{-1}\right]^2} \quad (20)$$

The integrals appearing here have been tabulated.[10]

Results.—A plot of $\sigma/(4a)$ versus ka is shown in the graph. This is a plot of equation (20). For small ka we have Rayleigh scattering. As ka increases the curve performs a damped oscillation about $\sigma/(4a) = 1$. And for $ka \to \infty$ it can be shown by means of asymptotic representations of Bessel functions and their integrals that $\sigma \to 4a$.

According to Babinet's principle[11] the problem we have discussed is complementary to the problem of scattering by a slit for a normally incident wave polar-

ized perpendicular to the axis of the slit. For intermediate values of ka our curve behaves qualitatively as Morse and Rubinstein's:[1] the quantitative agreement is less good, the deviation being due to our choice of an overly simplified trial function. It is quite remarkable how such a rough approximation of the induced current yields fairly good results over the entire spectrum.

* Most of the work discussed in this note was supported by the Office of Naval Research.

[1] Sieger, B., *Ann. d. Physik*, **27**, 622 (1908). Wien, W., *Jahresbericht d. deutsch Math. Verein*, **15**, 42 (1906). Schwarzschild, K., *Math. Ann.*, **55**, 177 (1902). Morse, P. M., and Rubinstein, P. J., *Phys. Rev.*, **54**, 895 (1938). Lord Rayleigh, *Phil. Mag.*, (5) **43**, 259 (1897); **44**, 28 (1897). Epstein, P. S., *Encyklopädie für die Mathematischen Wissenschafften*, Vol. 5, III.

[2] Miles, J. W., *J. of Math. and Phys.*, **28**, 223 (1950). Groschwitz and Hönl, H., *Z. Physik*, **131**, 305 (1952). Sommerfeld, A., *Vorlesungen über theoretische Physik*, Bd. IV, Optik, 39. Müller, Rolf and Westpfahl, Konrad, *Z. für Physik*, **134**, 245 (1953).

[3] Moullin, E. B., and Phillips, F. M., *Proc. Inst. Elec. Engrs.*, Part IV, **99**, 137 (1952). Morse and Rubinstein, *loc. cit.* (ref. 1).

[4] Levine, H., and Schwinger, J., *Comm. Pure and App. Math.*, **3**, 355 (1950). See also Baker B. B., and Copson, E. T., *The Mathematical Theory of Huygens' Principle*, 2nd ed., Oxford (1950), Chapter V, § 6.2; Marcuvitz, N., in the New York University Seminar Notes, "Recent Developments in the Theory of Wave Propagation," § III D, New York (1953); and Erdélyi, A., *Acc. d. Sci. Torino, Atti*, **87** (1952–53) for the scalar case.

[5] Fock, V., *J. of Phys. of U.S.S.R.*, **10**, 130 (1946).

[6] Bouwkamp, C. J., *Physica*, **12**, 467 (1946). Meixner, J., *Ann. d. Phys.*, (6) **6**, 2 (1949).

[7] Papas, C. H., *J. App. Phys.*, **21**, 318 (1950).

[8] Twersky, V., Research Report No. EM-54, Math. Res. Group, N. Y. U.

[9] Brillouin, L., *J. App. Phys.*, **20**, 1110 (1949).

[10] Lowan and Abramowitz, *J. Math. Phys.*, **22**, 2 (1943). Watson, G. N., *Bessel Functions*, Cambridge (1945).

[11] Epstein, P. S., *loc. cit.* ref. 1, p. 511. Baker, B. B., and Copson, E. T., *loc. cit.* ref. 4.

THE QUANTUM CORRECTION IN THE RADIATION BY ENERGETIC ACCELERATED ELECTRONS

BY JULIAN SCHWINGER

HARVARD UNIVERSITY

Communicated December 30, 1953

The radiation emitted by high energy electrons in a large scale magnetic field is an essentially classical phenomenon, and was so discussed in a paper[1] of the author. It was there remarked that the condition for quantum effects to be unimportant is that the momenta of the radiated quanta be small compared with the electron momentum. At very high electron energies, this requirement reads

$$\hbar\omega << E, \qquad \omega \sim \omega_0 (E/mc^2)^3, \tag{1}$$

where

$$\omega_0 = c/R = (eH/mc)(mc^2/E) \tag{2}$$

is the rotational frequency in the orbit determined by the magnetic field **H**. Hence

the magnitude of the quantum correction should be indicated by

$$\frac{\hbar\omega_0}{E}\left(\frac{E}{mc^2}\right)^3 = \frac{\hbar/mc}{R}\left(\frac{E}{mc^2}\right)^2$$

$$= \frac{(e\hbar/mc)H}{mc^2}\,\frac{E}{mc^2}, \tag{3}$$

which is extremely small for all existing or contemplated magnetic devices. Recently, there have been several attempts at a quantum mechanical calculation[2] or discussion[3] of this emission process, which have led to varying degrees of doubt concerning the validity of (3) as a measure of the quantum correction. To aid in settling this question, we shall present an explicit calculation of the quantum correction, to the first order in $\hbar$.

At this level of accuracy, the spin degrees of freedom play no role for unpolarized particles. With a suitable canonical transformation, the Dirac Hamiltonian can be replaced by the classical

$$H = \left[c^2\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + (mc^2)^2\right]^{1/2}, \tag{4}$$

for a field described by a vector potential (this includes the magnetic field and the radiation field). The probability for emitting a photon, during the time interval $T = t_1 - t_2$, into the range $(d\mathbf{k})$ about the propagation vector $\mathbf{k}$ and with polarization vector $\mathbf{e}$, is obtained as the absolute squared matrix element of the operator

$$\left[\frac{(d\mathbf{k})}{(2\pi)^3}\frac{2\pi}{\hbar\omega}\right]^{1/2} e\int_{t_2}^{t_1} dt e^{i\omega t}\,(\mathbf{e}\cdot\mathbf{v}(t), e^{-i\mathbf{k}\cdot\mathbf{r}(t)}), \tag{5}$$

where $\mathbf{v}(t)$ is the particle velocity, and $(\quad,\quad)$ indicates the symmetrized product of the operators. This matrix element refers to specific initial and final particle states. On summing over all final particle states, we obtain the probability for the radiation process as an expectation value computed for the initial particle state,

$$\frac{(d\mathbf{k})}{(2\pi)^3}\frac{2\pi e^2}{\hbar\omega}\left\langle\int_{t_2}^{t_1} dt\int_{t_2}^{t_1} dt' e^{-i\omega(t-t')}(\mathbf{e}\cdot\mathbf{v}(t), e^{i\mathbf{k}\cdot\mathbf{r}(t)})(\mathbf{e}\cdot\mathbf{v}(t'), e^{-i\mathbf{k}\cdot\mathbf{r}(t')})\right\rangle \tag{6}$$

If we are not concerned with the radiation of a specific polarization, one should replace $\mathbf{e}\cdot\mathbf{v}(t)\mathbf{e}\cdot\mathbf{v}(t')$ with $\mathbf{n}\times\mathbf{v}(t)\cdot\mathbf{n}\times\mathbf{v}(t') = \mathbf{v}(t)\cdot\mathbf{v}(t') - \mathbf{n}\cdot\mathbf{v}(t)\mathbf{n}\cdot\mathbf{v}(t')$, where $\mathbf{n}$ is the unit vector directed along $\mathbf{k}$. The relation

$$(i\mathbf{k}\cdot\mathbf{v}(t), e^{i\mathbf{k}\cdot\mathbf{r}(t)}) = (d/dt)e^{i\mathbf{k}\cdot\mathbf{r}(t)} \tag{7}$$

then enables us to write the probability for radiation, without regard to polarization, as

$$\frac{(d\mathbf{k})}{(2\pi)^3}\frac{2\pi e^2c^2}{\hbar\omega}\left\langle\int_{t_2}^{t_1} dt\int_{t_2}^{t_1} dt' e^{-i\omega(t-t')}\left[\frac{1}{c^2}(\mathbf{v}(t), e^{i\mathbf{k}\cdot\mathbf{r}(t)})\cdot(\mathbf{v}(t'), e^{-i\mathbf{k}\cdot\mathbf{r}(t')}) - e^{i\mathbf{k}\cdot\mathbf{r}(t)}e^{-i\mathbf{k}\cdot\mathbf{r}(t')}\right]\right\rangle. \tag{8}$$

For a sufficiently extended time interval T, we can introduce new time variables $t \rightarrow t + {}^1/_2\tau$, $t' \rightarrow t - {}^1/_2\tau$, and regard the t integrand as the probability per unit time of the event. On multiplication with $\hbar\omega$ we find that the average power radiated into a unit solid angle enclosing $\mathbf{n}$, and contained in a unit frequency interval about ω, is

$$P(\mathbf{n}, \omega, t) = \frac{e^2}{4\pi^2}\frac{\omega^2}{c}\Big\langle\int_{-\infty}^{\infty} d\tau e^{-i\omega\tau}\Big[\frac{1}{c^2}\Big(\mathbf{v}\Big(t+\frac{1}{2}\tau\Big), e^{i\mathbf{k}\cdot\mathbf{r}(t+{}^1/_2\tau)}\Big)\cdot\Big(\mathbf{v}\Big(t-\frac{1}{2}\tau\Big), e^{-i\mathbf{k}\cdot\mathbf{r}(t-{}^1/_2\tau)}\Big) - e^{i\mathbf{k}\cdot\mathbf{r}(t+{}^1/_2\tau)}\, e^{-i\mathbf{k}\cdot\mathbf{r}(t-{}^1/_2\tau)}\Big]\Big\rangle. \quad (9)$$

If the order of operator multiplication is not significant, this result is equivalent to the classical formula (I.38) of reference 1.

We shall now investigate the commutation properties of the operators $\mathbf{r}(t)$ and $\mathbf{v}(t)$, to the first order in $\hbar$. Thus, the equation

$$(\mathbf{v}, H) = \Big(\mathbf{p} - \frac{e}{c}\mathbf{A}\Big)c^2 \quad (10)$$

is solved, to this accuracy, by

$$\mathbf{v} = \Big(H^{-1}, \Big(\mathbf{p} - \frac{e}{c}\mathbf{A}\Big)c^2\Big), \quad (11)$$

from which we derive

$$[r_i, v_j] = i\hbar\frac{\partial v_j}{\partial p_i} = i\hbar\frac{c^2}{E}\Big[\delta_{ij} - \frac{1}{c^2}v_i v_j\Big]. \quad (12)$$

Consistent with the approximation program, we have on the right replaced the operator H by the classical determinate energy E, and disregarded the non-commutativity of v_i and v_j. The latter is described, to the same approximation, by

$$\frac{1}{c^2}\mathbf{v}\times\mathbf{v} = i\hbar\frac{ec}{E^2}\Big[(1 - v^2/c^2)\mathbf{H} + \frac{1}{c^2}\mathbf{v}\mathbf{v}\cdot\mathbf{H}\Big]. \quad (13)$$

The classical motion occurs in the plane perpendicular to the magnetic field, $\mathbf{v}\cdot\mathbf{H} = 0$, so that (13) implies the following uncertainty principle for the velocity components in this plane,

$$\frac{1}{c^2}\Delta v_1\Delta v_2 \geq \frac{1}{2}\hbar\frac{ecH}{E^2}(1 - v^2/c^2), \quad (14)$$

from which we infer

$$\frac{1}{c^2}[(\Delta v_1)^2 + (\Delta v_2)^2] \geq \hbar\frac{ecH}{E^2}(1 - v^2/c^2) = \frac{\hbar\omega_0}{E}(1 - v^2/c^2). \quad (15)$$

Since the particle speed occurs significantly only in the combination $1 - v^2/c^2$, the effect of the non-simultaneous specifiability of v_1 and v_2 is measured by the coefficient $\hbar\omega_0/E = e\hbar cH/E^2$, which is negligible compared to (3), and decreases

with increasing energy. Hence, at sufficiently high energies, the components of $\mathbf{v}$ are determinate quantities.

The symmetrized products of $\mathbf{v}$ and $e^{\pm i\mathbf{k}\cdot\mathbf{r}}$ can be replaced by

$$(\mathbf{v}, e^{i\mathbf{k}\cdot\mathbf{r}}) = \left[\mathbf{v} - \frac{\hbar c^2}{2E}\left(\mathbf{k} - \frac{1}{c^2}\,\mathbf{v}\mathbf{k}\cdot\mathbf{v}\right)\right]e^{i\mathbf{k}\cdot\mathbf{r}} \tag{16}$$

and

$$(\mathbf{v}, e^{-i\mathbf{k}\cdot\mathbf{r}}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\left[\mathbf{v} - \frac{\hbar c^2}{2E}\left(k - \frac{1}{c^2}\,\mathbf{v}\mathbf{k}\cdot\mathbf{v}\right)\right]. \tag{17}$$

We then combine the contiguous quantities, $e^{i\mathbf{k}\cdot\mathbf{r}(t+1/2\tau)}$ and $e^{-i\mathbf{k}\cdot\mathbf{r}(t-1/2\tau)}$, with the aid of the theorem

$$e^A e^B = e^{A+B}e^{1/2[A,B]}, \tag{18}$$

which is exact if [A, B] commutes with A and B, and suffices for the purposes of the approximation we are employing. Since the values of τ that contribute to the integral (9) are such that $\omega\tau(1 - v^2/c^2) \sim \omega_0\tau(1 - v^2/c^2)^{-1/2} \sim 1$, we have $\omega_0\tau << 1$, and one can use the expansion

$$\mathbf{r}\left(t \pm \frac{1}{2}\tau\right) = \mathbf{r}(t) \pm \frac{1}{2}\tau\mathbf{v}(t) + \frac{1}{8}\tau^2\dot{\mathbf{v}}(t) \pm \frac{1}{48}\tau^3\ddot{\mathbf{v}}(t). \tag{19}$$

Thus

$$\begin{aligned}\mathbf{r}\left(t + \frac{1}{2}\tau\right) - \mathbf{r}\left(t - \frac{1}{2}\tau\right) &= \tau\mathbf{v}(t) + \frac{1}{24}\tau^3\ddot{\mathbf{v}}(t)\\ &= \left(\tau - \frac{1}{24}\,\omega_0{}^2\tau^3\right)\mathbf{v}(t),\end{aligned} \tag{20}$$

and

$$\begin{aligned}\left[\mathbf{k}\cdot\mathbf{r}\left(t + \frac{1}{2}\tau\right), \mathbf{k}\cdot\mathbf{r}\left(t - \frac{1}{2}\tau\right)\right] &= -\tau[\mathbf{k}.\mathbf{r}, \mathbf{k}.\mathbf{v}] - \frac{1}{24}\tau^3[\mathbf{k}.\mathbf{r}, \mathbf{k}.\ddot{\mathbf{v}}]\\ &= -i\hbar\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\right)\left[\left(\tau - \frac{1}{24}\,\omega_0{}^2\tau^3\right)\mathbf{k}\cdot\mathbf{v}\right]\\ &= -2i\frac{\hbar\omega}{E}\,\omega\left[\left(1 - \frac{1}{c}\,\mathbf{n}\cdot\mathbf{v}\right)\tau + \frac{1}{24}\,\omega_0{}^2\tau^3\right],\end{aligned} \tag{21}$$

in which we have omitted commutators containing only velocity components, and introduced simplifications appropriate to the nature of the angular distribution, $\frac{1}{c}\mathbf{n}\cdot\mathbf{v} \sim 1$. We should also note that the velocity factors multiplying the exponentials in (9) combine into

$$\left(\frac{1}{c^2}\,\mathbf{v}\left(t + \frac{1}{2}\tau\right)\cdot\mathbf{v}\left(t - \frac{1}{2}\tau\right) - 1\right)\left(1 + \frac{\hbar\omega}{E}\,\frac{1}{2c}\,\mathbf{n}\cdot\left(\mathbf{v}\left(t + \frac{1}{2}\tau\right) + \mathbf{v}\left(t - \frac{1}{2}\tau\right)\right)\right) \tag{22}$$

where the second factor can be approximated by $1 + (\hbar\omega/E)$. The resulting form of (9) is

$$P(\mathbf{n}, \omega, t) = \frac{e^2}{4\pi^2}\frac{\omega^2}{c}\left(1 + \frac{\hbar\omega}{E}\right)\int_{-\infty}^{\infty} d\tau\left(\frac{1}{c^2}\mathbf{v}\left(t + \frac{1}{2}\tau\right)\cdot\mathbf{v}\left(t - \frac{1}{2}\tau\right) - 1\right)$$
$$\cdot\exp\left[-i\omega\left(1 + \frac{\hbar\omega}{E}\right)\left(\left(1 - \frac{1}{c}\mathbf{n}\cdot\mathbf{v}\right)\tau + \frac{1}{24}\omega_0{}^2\tau^3\right)\right]. \quad (23)$$

We have thus obtained the simple result that first order quantum effects are included by making the substitution

$$\omega \rightarrow \omega\left(1 + \frac{\hbar\omega}{E}\right) \quad (24)$$

in the classical formula for $\omega^{-1}P(\mathbf{n}, \omega, t)$. The same statement applies to $\omega^{-1}P(\omega, t)$, since the correction factor does not depend upon the emission direction. On referring to the classical formula (II.16) of reference 1, we see that the power radiated per unit frequency range about ω is given by

$$P(\omega, t) = \frac{3^{3/2}}{4\pi}\frac{e^2}{R}\left(\frac{E}{mc^2}\right)^4\frac{\omega_0\omega}{\omega_c{}^2}\int_{\omega/\omega_c(1+(\hbar\omega/E))}^{\infty} d\eta K_{5/3}(\eta), \quad (25)$$

where

$$\omega_c = \frac{3}{2}\omega_0(E/mc^2)^3. \quad (26)$$

Hence, the first order quantum correction is

$$\delta P(\omega, t) = -\frac{3^{3/2}}{4\pi}\frac{e^2}{R}\left(\frac{E}{mc^2}\right)^4\frac{\omega_0}{\omega_c}\left(\frac{\omega}{\omega_c}\right)^3\frac{\hbar\omega_c}{E}K_{5/3}(\omega/\omega_c), \quad (27)$$

and the corresponding correction to the total power emerges as

$$\delta P(t) = -3^{1/2}\frac{55}{36}\omega_0\frac{e^2}{R}\left(\frac{E}{mc^2}\right)^4\frac{\hbar\omega_c}{E}, \quad (28)$$

with the aid of the integral

$$\int_0^{\infty} d\xi\, \xi^3 K_{5/3}(\xi) = \frac{55}{27}\pi. \quad (29)$$

It is thus found that, with regard to the total radiated power, the effect of first order quantum corrections is to multiply the classical value, $\frac{2}{3}\omega_0(e^2/R)(E/mc^2)^4$, by the factor[4]

$$1 - 3^{1/2}\frac{55}{16}\frac{\hbar/mc}{R}\left(\frac{E}{mc^2}\right)^2. \quad (30)$$

[1] Schwinger, J., *Phys. Rev.*, **75**, 1912 (1949).

[2] Parzen, G., *Ibid.*, **84**, 235 (1951) (corrected by Judd, Lepore, Ruderman, and Wolff, *Ibid.*, **86**, 123 (1952), and Olsen, H., and Wergeland, H., *Ibid.*, **86**, 123 (1952)). Neuman, M., *Ibid.*, **90**, 682 (1953).

[3] Schiff, L. I., *Am. J. Phys.*, **20**, 474 (1952).

[4] While writing this paper, the author was delighted to encounter, in the December, 1953, issue of *Physics Abstracts*, a report of a paper by Sokolov, A. A., Klepikov, N. P., and Ternov, I. M., *Doklady Akad. Nauk SSSR*, **89**, 665 (1953), in which the identical result is quoted as obtained by a calculation based on the Dirac equation.

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

REFERENCES to literature, numbered consecutively, will be placed at the end of the article and short footnotes should be avoided. It is suggested that references to periodicals be furnished in some detail and in general in accordance with the standard adopted for the Subject Catalogue of the International Catalog of Scientific Literature, viz., name of author, with initials following (ordinarily omitting title of paper), abbreviated name of journal, with place of publication, series (if any), volume, inclusive pages, year. For example: Montgomery, T. H., *J. Morph.*, Boston, **22**, 731–815 (1911); or, Wheeler, W. M., *Konigsberg, Schr. physik. Ges.*, **25**, 1–142 (1914).

MANUSCRIPTS will be accepted only from members of the Academy or from chairmen of the divisions of the National Research Council who will assume responsibility for the propriety and scientific standards of the paper and should be addressed to *Proceedings of the* NATIONAL ACADEMY OF SCIENCES, Attention Miss Mary D. Alexander, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois. Articles of which a member of the Academy is author or co-author may extend to 10 printed pages, those by non-members to 5 printed pages; space above these limits will be billed to the author at $21.00 per page. Accompanying the manuscript should be a statement by the author or his representative that he assumes responsibility for the charges if incurred.

GALLEY PROOF will be sent, with the understanding that charges for author's corrections shall be billed to him. Author should therefore make final revisions on the typewritten manuscripts. Page proofs will not be sent.

REPRINTS should be ordered when proof is returned, on the reprint order form which will be inclosed with the proof. The order blank gives a tabular estimation of the cost of reprints; a copy will be supplied at any time upon request.

CONTENTS

Page

BACTERIOLOGY.—CYCLICAL BEHAVIOR IN PNEUMOCOCCAL GROWTH AND TRANSFORMABILITY OCCASIONED BY ENVIRONMENTAL CHANGES . . . . . *By Rollin D. Hotchkiss* 49

BACTERIOLOGY.—DOUBLE MARKER TRANSFORMATIONS AS EVIDENCE OF LINKED FACTORS IN DESOXYRIBONUCLEATE TRANSFORMING AGENTS . . . . . *By Rollin D. Hotchkiss and J. Marmur* 55

BIOCHEMISTRY.—THE ISOLATION AND IDENTIFICATION OF THE MAJOR ANION FRACTION OF THE AXOPLASM OF SQUID GIANT NERVE FIBERS . . . . . *By Bernard A. Koechlin* 60

CHEMISTRY.—DIELECTRIC INCREMENTS IN AQUEOUS SOLUTIONS OF SYNTHETIC POLYELECTROLYTES . . . . . *By Howard M. Dintzis, J. L. Oncley, and Raymond M. Fuoss* 62

CHEMISTRY.—THE RATES OF REACTION BETWEEN ALKYL HALIDES AND TERTIARY AMINES AS DETERMINED CONDUCTOMETRICALLY . *By Robert P. Larsen and Charles A. Kraus* 70

ENGINEERING.—SONIC POINT ON SUPERSONIC AIRFOILS . . . . . *By T. Y. Thomas* 76

ENGINEERING.—REAR SHOCK AND PRESSURE ON SUPERSONIC AIRFOILS . . . . . *By W. F. Brown and T. Y. Thomas* 83

GENETICS.—GENETIC CHANGES IN AMERICAN POPULATIONS OF *Drosophila Melanogaster* . . . . . *By P. T. Ives* 87

GENETICS.—PSEUDO-ALLELISM AT THE VERMILION LOCUS IN *Drosophila Melanogaster* . . . . . *By M. M. Green* 92

GEOPHYSICS.—NUMERICAL PREDICTION OF CYCLOGENESIS . . . . . *By J. G. Charney* 99

MATHEMATICS.—ARITHMETIC GENERA AND THE THEOREM OF RIEMANN-ROCH FOR ALGEBRAIC VARIETIES . . . . . *By Friedrich Hirzebruch* 110

MATHEMATICS.—ON THE CONTINUOUS GOLD-MINING EQUATION . . . . . *By Richard Bellman and Sherman Lehman* 115

MATHEMATICS.—ON A PARTICULAR TYPE OF CONVERGENCE TO A SINGULAR MATRIX . . . . . *By E. Inonu and E. P. Wigner* 119

MATHEMATICS.—ON NON-NEGATIVE-VALUED MATRICES . . . . . *By Y. K. Wong* 121

PHYSICS.—SOME RELATIONS FOR CRYSTALS WITH SUBSTRUCTURES . . *By M. J. Buerger* 125

PHYSICS.—ON DIFFRACTION BY A STRIP . . . . . *By A. Erdélyi and C. H. Papas* 128

PHYSICS.—THE QUANTUM CORRECTION IN THE RADIATION BY ENERGETIC ACCELERATED ELECTRONS . . . . . *By Julian Schwinger* 132